UNCLASSIFIED
Security Classification

AD748407

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* U. S. Army Research Office-Durham Box CM, Duke Station Durham, North Carolina 27706	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

PROCEEDINGS OF THE 1972 U. S. ARMY OPERATIONS RESEARCH SYMPOSIUM, "RISK ANALYSIS".

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(First name, middle initial, last name)*

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August 1972		

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	NONE
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
NONE	Office of the Chief of Research and Development, Department of the Army

13. ABSTRACT

This is a technical report containing papers presented in person and by title at the Eleventh Annual U. S. Army Operations Research Symposium. The basic theme of the symposium was "Risk Analysis".

14. Key Words

Applied Decision Theory
Data Analysis
Decision Making
Decision Theory
Measures of Effectiveness
Operations Research
Risk Analysis

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

UNCLASSIFIED
Security Classification

PROCEEDINGS

of the

United States Army

OPERATIONS RESEARCH SYMPOSIUM

15-18 May 1972

Sponsored by

Office, Chief of Research and Development

Department of the Army

DDC
RECEIVED
SEP 18 1972
A

Hosted and Conducted by

U.S. Army Research Office-Durham

Durham, North Carolina

11

This document has been approved for public release and sale; its distribution is unlimited.

U. S. ARMY OPERATIONS RESEARCH SYMPOSIUM

15-18 May 1972

FOREWORD

The Eleventh Annual U. S. Army Operations Research Symposium was held on 15-18 May 1972. These symposia, normally held in the spring of each year, are sponsored by the Office of the Chief of Research and Development, and conducted by the U. S. Army Research Office-Durham.

This year some new aspects were introduced, such as, expository presentations on decision analysis, panel sessions on related subjects and a considerable amount of audience participation. The response of the participants is being evaluated to determine the future structure of the symposia.

This volume contains invited and contributed papers and major addresses. Some of the presentations at the symposium are not included here either because the paper was not formalized or the speaker chose not to have his remarks published.

The technical program for the symposium was planned and organized by Mr. Jerome H. N. Selman of the U. S. Army Munitions Command and ARO-D. This office is indebted to Mr. Selman for his outstanding efforts on our behalf. We also appreciate the valuable assistance of those who organized and participated in the various sessions and panels of our symposium.

Lothrop Mittenthal

LOTHROP MITTENTHAL
COL, GS
Commanding
U. S. Army Research Office-Durham

TABLE OF CONTENTS

WELCOME

by

COLONEL LOTHROP MITTENTHAL
CO, U. S. Army Research Office-Durham

On behalf of the Army Research Office here in Durham I'd like to welcome you all here. Perhaps some of you should be welcoming me because many of you have been here a number of times and I have never been to one of these before. Nevertheless I would like to say a few words about this meeting and some of the other things we do here. As you are probably well aware, this has turned out to be a very popular symposium. This is the eleventh in the series and in fact, unfortunately, because of space limitations, again we had to turn some people away. Our function here in Durham, of course, is considerably broader than this. Our main mission is supporting basic research in the physical and engineering sciences and the environmental sciences through grants and contracts to universities. In addition, we do a number of other things but perhaps we're second most famous for our scientific services; that is, the consultant services we have to Army laboratories which provides experts on short notice. A third function that we have in this office is supporting scientific symposia, of which this is perhaps our most prominent one. There are some twenty-three others this year that we are supporting in various places. A fourth function, which is somewhat related to this but a little more "junior" category, is that we support and sponsor the National Junior Science and Humanities Symposium in which we have thirty-one regional symposia of high school students who present their own scientific papers - people a bit junior to you in scientific status and age - and then they convene annually for a national symposium. Those are our major functions.

Some of you who are familiar with this office may wonder about our status. There was some indication we would be moving from here. We are not going to be leaving the Duke campus for another few years. We must leave here by February, 1975, but it has been decided to stay here until our lease expires. As you also know, Duke University has chosen to sever their contractual relationship with us. That is now being phased out and we have a new contract for scientific services with Battelle-Columbus Laboratories, which on May 1, set up an office here in Durham. We trust we will be able to give you just as good service through Battelle as we were able to give you through Duke.

Some of you also may have heard about student unrest here. Of course, you don't have to worry about it too much right now. Final exams are over and graduation was Sunday and only the most eager students could remain behind after that to harass us here. As far as I know, none have. We have been asked by students to leave the campus on the grounds that we are imperialist warmongers. We couldn't really agree with them and declined to do so. That's really been about the net effect of it, and I doubt very much that our meeting will be in any way impeded by that. I hope you will have a stimulating symposium.

EULOGY FOR DR. GEORGE E. NICHOLSON, JR.

by

DR. MARION R. BRYSON
U. S. Army Combat Developments Command
Systems Analysis Group
Fort Belvoir, Virginia

Dr. George E. Nicholson, Jr. was a man of unusual stature in the scientific world. His reputation as an outstanding leader in both statistics and Operations Research was international.

Dr. Nicholson was born on June 21, 1919 in Brooklyn, New York and obtained his early education in that city and at the University of North Carolina. Being unable to enlist in the Armed Forces during World War II, he joined the war effort at Columbia University as a mathematician with the Columbia project. There he became interested in the emerging field of Operations Analysis. He served with distinction as a civilian operations analyst for the Air Force in Saipan until the end of the war.

In 1946, he began a long career at the University of North Carolina. He earned a Ph.D. in statistics at UNC in 1948 and was immediately appointed to the faculty there. In 1952, he was appointed chairman of the Department of Statistics at North Carolina; a position he held for the following 19 years.

Dr. Nicholson's versatility led to his service on many important tasks throughout his career. These include service as a U. S. representative to Europe on a Weapons Evaluation Committee in NATO in 1964 and a similar position in Japan in 1965. For many years he headed an Air Force Operations Research unit in Chapel Hill.

His many honors include the Medal of Freedom for his Air Force work in Saipan, the Department of Defense Exceptional Civilian Service Medal for his NATO work, Fellowship in the American Statistical Association, and Fellowship in the Institute of Mathematical Statistics, an organization he served as secretary for many years.

Of particular interest to those honoring Dr. Nicholson here today is his work in the Army Operations Research symposia. He was a leader, advisor, and willing participant in the planning and conduct of these meetings from their beginning in 1962. In the first two symposia he was a session chairman. In 1964 he was honored by the Symposium Planning Committee with his appointment as general chairman of the meeting, an honor repeated in 1966. In 1965 he was the closing speaker when he delivered the critique address. His guidance and counsel during the succeeding years served to insure the success of the AORS. For the 1972 symposium, the current one, he was appointed program chairman with overall responsibility for the organization of the technical sessions. What you will hear in the next three days was very much influenced by him in the formative months when wise guidance is so necessary.

Preceding page blank

Nick never met a stranger. His outgoing personality and friendly nature made each of us who knew him feel as if we were one of his special friends; as, indeed, we were. Although his assistance was widely sought, he always had time to help when called upon. Nick took a personal interest in the problems of those around him. He was no less willing to talk at length with the beginning student than he was with the internationally famous. Nick received his greatest pleasure from his family, his friends, his colleagues; but he always gave more than he received.

With the death of Dr. George E. Nicholson, Jr., on 3 December 1971, the scientific community lost a leader, military Operations Research lost a founding father, but we, his many friends, had our sorrow tempered with thankfulness for having shared in the life of a great man.

INTRODUCTION OF KEYNOTE SPEAKER

by

BRIGADIER GENERAL CHARLES D. DANIEL, JR.
Director of Army Research

I'd like to add my welcome to you to that of Colonel Mittenthal for it is my distinct pleasure to introduce our Keynote Speaker, Mr. Abraham Golub, Scientific Advisor to the Assistant Chief of Staff for Force Development, United States Army.

The purpose of the symposia is to provide a broad forum for the exchange of information and ideas concerning the uses and techniques of Operations Research and to focus on the needs and accomplishments of the Army in this area. One of the primary uses of Operations Research in the Army is in force development. This activity determines to a large extent how the Army will be organized, equipped and trained to accomplish its many missions. There will never be enough resources, of course, to perform every single task believed necessary for defense. It is the job of the analyst to make detailed examinations to insure the most efficient allocation of the available resources. Thus, the job of force development and the resultant effectiveness of the Army depends upon how well the analysts and managers at all levels conduct those examinations. To provide the Army with qualified analysts and managers, ACSFOR is deeply involved in the Operations Research Analysis Officer Career Program. This includes assistance to the Deputy Chief of Staff for Personnel in the formulation of policy and determination of standards of selection of program members and in the designation of ORSA specialists and executive positions within the Army. The Keynote Speaker of this year's symposium, Mr. Abraham Golub, occupies a key position in the activity that I have just described. As Scientific Advisor to the Assistant Chief of Staff for Force Development, Mr. Golub advises General Williams, who has general staff responsibility for the development and implementation of Army force development plans and the development of the requirements documents which lead to the acquisition of specific items of equipment. To his current position, Mr. Golub brought many years of experience in Operations Research and I would like to give you a few of the highlights of his career.

Mr. Golub received his B. A. in Mathematics from Brooklyn College. In 1942, he joined the Office of Chief of Ordnance and was called to active duty from the period 1943 to 1946. Upon his release from service, he joined the Ballistics Research Laboratory and, in succeeding positions of increasing responsibility, was appointed the Associate Technical Director of the laboratory in 1962. During his period of service at BRL, Mr. Golub received his M. A. in Mathematical Statistics at the University of Delaware in 1949, and he taught his specialty there from the period 1952 to 1954. He continued his graduate studies at George Washington University from the period 1954 to 1959. In 1964, he was appointed Deputy Special Assistant for Operations Research to the Assistant Secretary of the Army for Financial Management and for the period 1966 to 1969 he served as Assistant Deputy

Under Secretary of the Army for Operations Research. In 1969, he accepted his current position in ACSFOR. Mr. Golub is currently a member of the Operations Research Society of America, the American Mathematical Society and the American Ordnance Association, among others. He has published several articles, one of which bears the title, "For Better Use of Present ORSA Talent." Today, Mr. Golub will speak to you on the subject of "Risk Analysis Planning for Today's Army." Since the theme of this symposium is "Risk Analysis", Mr. Golub's title suggests a challenge to the participants to apply this concept to its fullest potential, not only in our attempts to solve today's problems but also in our preparation for the development of future forces. It's a very great privilege and a pleasure to introduce our distinguished keynoter, Mr. Golub.

"RISK ANALYSIS AND PLANNING FOR TOMORROW'S ARMY"*

by

DR. ALRAHAM GOLUB
Office of the Assistant Chief
of Staff for Force Development

Thank you, General Daniel for that gracious introduction. But first let me put things straight for the record. As many of you know, my boss General Williams, The Assistant Chief of Staff for Force Development, was to have given this address. Unfortunately, he couldn't attend--for fairly obvious reasons these days--and he asked me to convey his disappointment that he could not be with you. He also asked me--no, directed me--to fill in for him.

Now even though I was informed of my participation in these sessions only four days ago, I did consider two possible options for accomplishing my mission. First, I could simply relate some of the material previously prepared by General Williams and express his views--or I could prepare a paper from scratch and present some of my own views. In making that decision I examined my most recent risk analysis to see whether they were sufficiently successful to warrant discussion before this body. However, just as I began to give these two options serious consideration, General Williams called me again to make sure that I would convey his views and, above all, that I would make it short--so that you people could get down to the business of your meetings. So that's what I will do.

Incidentally, I did examine my most recent risk analysis for evidence of success, and as it turned out General Williams was quite right in his choice of my option. Let me explain.

Some time ago my wife indicated that she would like to have our son and daughter come home from school to help me celebrate my birthday. She thought it would be a good idea to have the family spend that day together. The following day I checked my schedule, exercised a measure of control, moved some meetings around and assured my wife that there was little question (probability close to 1) but that I would be home that day and to go ahead and notify the kids. Well, I can now report three facts that relate importantly to the success of my analysis. First my daughter arrived home as scheduled; second my son arrived home as scheduled and third my birthday is May 16. As you can see, General Williams was undoubtedly right in his guidance to me.

The theme you have selected for this year's meeting, "Risk Analysis," is most timely and opportune because of the rapidly changing environment we are living in today. From the standpoint of the ACSFOR I can summarize this new environment with a few terse phrases: Reduced Manpower, Higher Personnel Costs, Lower Budgets, Shifting National Priorities, and--

*Keynote Address, United States Army Operations Research Symposium, Durham, North Carolina, 16 May 1972.

perhaps most difficult of all--Unprecedented Public Scrutiny of Our Defense Acquisition Processes....I'm sure you are all aware of these factors. My point is that every one of these factors contributes added pressure to perform the force development and equipment acquisition processes as skillfully as possible. It is not through lack of honest desire or of trying--we invent and invoke new sets of controls and techniques when we think it necessary--but all too often we seem to be frustrated in our attempts to overcome the major cost overruns, schedule slippages and hardware deficiencies that have plagued us for so long. It almost seems as though there has been something fundamentally wrong or lacking in our practices, and General Williams has become convinced that the early and explicit consideration of risk may be an important essential that is lacking.

Before I get into the reasons why the ACSFOR is interested in risk analysis, and the ways in which he hopes it can provide assistance to the difficult job of force development, I'd like to take a few minutes to briefly explore how he views risk and "Risk Analysis". Risks in themselves are nothing new. The military has always lived in an atmosphere of risks--both in peacetime planning and in actual war--and they have always tried to analyze risks. I'm sure the military's oft cited "Calculated Risk" is a term you are all familiar with. What impresses General Williams is that the analysis of risk is being formalized and analytical techniques are being developed to help the military 'Calculate' the "Calculated Risk". There seem to be a lot of different terms around to describe this kind of activity, and the ACSFOR was both amused and a little perplexed at the number of variations on the theme of "Risk" that do exist. Let me show you what he collected from just two pages of some background material. He believes that these are really all names for the same kind of work, and that a good deal of such work has been done previously under different names. Generally, the ACSFOR regards risk simply as meaning the "Probability of failure (or success)", but perhaps more important, he always thinks in terms of the "Impact of that Failure". For example, if someone should mention the risks associated with the Army's Advanced Attack Helicopter Program, I think General Williams would think first and foremost in terms of the impact should that program fail. He believes that if the analyst is fully aware of the impact of a possible failure he will most likely come up with a more careful and better risk analysis. I think you'll understand this better when I describe the responsibilities of the ACSFOR.

When Secretary Packard first gave prominence to the term "Formal Risk Analysis", he was addressing primarily the problems associated with the weapon system acquisition process. The ACSFOR's generalized impression of risk analysis is, I believe, consistent with this. He views it as a process wherein the risks associated with a particular developmental program are identified and evaluated, and alternative courses of action for reducing risk are generated. He also regards it as a continuing and iterative type of process rather than a one-time effort. I said I would "briefly explore" the ACSFOR's view of risk analysis, so I'll keep my word and stop now. (Figure 1)

VARIOUS CONSIDERATIONS OF RISK

RISK

RISK ATTITUDE

RISK PREFERENCE

RISK POLICY

RISK ASSESSMENT

RISK MANAGEMENT

RISK ANALYSIS

FIGURE 1.

General Williams is aware also that we are beginning to see far greater concern with risk and uncertainties at all levels of the national defense effort. At one end of the spectrum we see attempts to perform the kinds of risk analyses that I have been describing. For example, there have been recent publications which are clearly in line with the highly focused technical evaluatiors intended by Secretary Packard and which address specific systems such as the Lance Missile and the new 105MM Towed Howitzer. He applauds such analytical efforts, for such analyses can be the key building blocks in evaluating the uncertainties associated with more aggregated systems. Moreover, he recognizes how these risk analyses interface with <u>Decision Analysis</u>: The process of sorting out the best combination of alternatives.

At the other end of the spectrum we now have what Secretary Laird has referred to as "Net Assessment". This is a kind of risk analysis at the national level in which all factors--Military, Technological, Political and Economic--are examined to see which factors impede and which enhance the achievement of our national security objectives. In these assessments we weigh the capabilities of potential enemies against our own capabilities and those of our allies. Out of this comparison comes a balance or net effect which is one measure of the risk associated with achieving a particular objective.

Let me now take some time to explain how ACSFOR fits into the Army's scheme of things--and show why he has a special interest in the potential benefits of risk analysis.

General Williams hasn't always found it easy to explain to others what it is that The Assistant Chief of Staff for Force Development is supposed to do. The title isn't quite as intuitively or as historically appreciated as--for example--The Assistant Chief of Staff for Intelligence, or the Deputy Chief of Staff for Personnel. Not too long ago he was attempting to describe the job to a British General Officer. After a minute or so the British Officer interrupted him to say, "Oh, now I know what it is you do--we have essentially the same job here, but we call it The Assistant Chief of Staff for Size and Shape". Acutally, that's not too far off the mark. General Williams frequently describes his job by analogy with a weaver. He's the fellow with the loom there in the middle--in the middle of The Army Staff and in the middle of The Force Development Process. Now let's look at what's going on in all this weaving. If there's any one thing that you can say about ACSFOR, it is that the ACSFOR's job is to determine the requirement for people and for things to man the force structure. Now how does this come about? ACSI develops the threat forecast and then DCSOPS working closely with The Chief of Staff in the joint arena establishes the strategic concepts and determines the major units required to carry out the joint strategy. ACSFOR rounds out the force structure by way of The Army Force Development Plan which establishes the requirement. Once ACSFOR identifies this requirement, everyone else has a responsibility for doing something. DCSPER gets the men and is responsible with CONARC for their training. DCSLOG with AMC gets the equipment. CDC comes up with ideas, concepts, and force designs: CRD determines development effort required. ACSC-E sets the policy and standards for Army communication.

When all this effort is brought together by the ACSFOR you see the Army force structure represented by SCARF which is the end product of the ACSFOR. The ACSFOR is the integrator of all staff activities. (Figure 2)

If General Williams were pressed to give a one sentence summary of the mission of the ACSFOR, I think he would say this: (Figure 3)

There is, of course, a great deal implied by that simple statement. For example, it means properly trained, fully equipped, adequately supported, continually modernized, properly organized, and so forth. Fulfillment of all these implied goals is what adds the extra dimension of challenge to the ACSFOR's job. As you might expect, this job often takes on the aspects of crystal-ball gazing and juggling.

Among the things the ACSFOR must try to foresee are: Congressionally-imposed manpower ceilings, levels and trends of future Army budgets, civilian and military pay scales, dates-quantities-and-capabilities of modernization equipment, future OSD obligations such as special mission forces, and finally--possible changes in the Army's strategic requirements. Clearly, there is moderate to substantial uncertainty in all these factors, and a pretty good risk of being wrong in several areas.

The Defense Department and the Army have a systematic procedure for developing and updating the Army program each year. It all begins with preparation of the force development plan by the ACSFOR's office. One of the realities we must cope with is that each year we must begin with the army structure that is in-being. We begin with what we have.

Another reality is that there is a substantial fixed base in this Army over which we have only partial control. This fixed base requirement includes the posts, camps and stations that we operate--including the Pentagon contingent--the army school system which we must maintain, and the special missions which are assigned to us by the Office of the Secretary of Defense in the national effort: for example, The Safeguard System.

Currently, a good approximation of this fixed base requirement requires about 50 percent of the Army budget. In effect, this much must be spent whether we have division mission forces to meet our contingencies or not. The other 50 percent must provide not only for maintaining the division mission forces with adequate equipment and supplies but also for the cost of reserve components, the annual cost of military construction, and the amount of money it takes to develop and introduce new weapon systems and other equipment.

What all this reduces to is that we are given a budget (or rather half a budget) and a statement of strategy requirements, and we are directed to mold a force that is consistent with both the budget constraint and the Army's mission within DOD's total force planning concept. Although we pay particular attention to the fiscal year immediately ahead, we also contribute to the annual update of the five year defense program. Thus, we are constantly striving to develop a program and an Army force structure which meets both the specifications of national strategy and the financial resources available.

OACSFOR AS A WEAVER

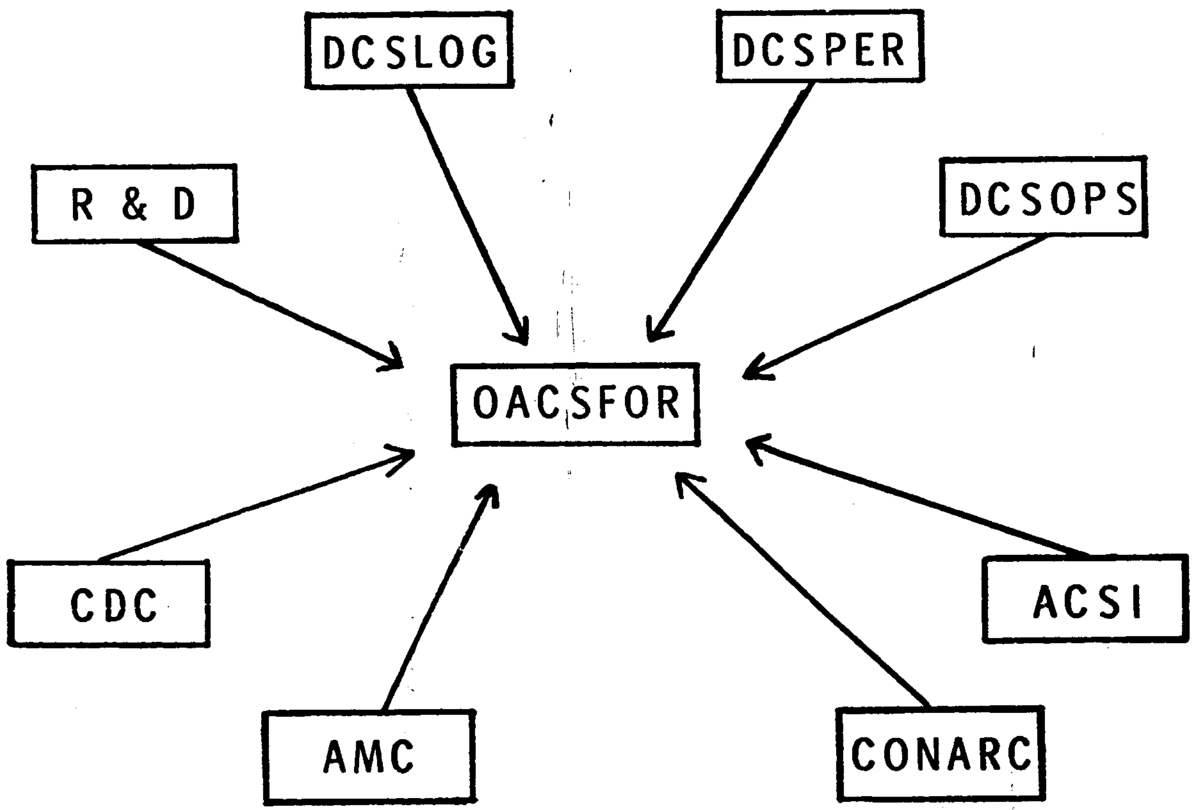


FIGURE 2.

THE ACSFOR'S MISSION

"TO PROVIDE THE TRAINED, EQUIPPED, AND SUPPORTED UNITS WHICH ARE REQUIRED IN OUR FORCE STRUCTURE TO ACCOMPLISH THE WORLD-WIDE MISSIONS OF THE ARMY."

FIGURE 3.

To summarize my remarks on the responsibilities of the ACSFOR: He is the manager of the force structure in being and the developer of a force structure which in the future will be equipped with the best weapons and equipment which can be made available, will contain well-trained combat effective units, and will be organized in such a manner that combat forces can be sustained in a theater of operations.

Even from this brief description of the ACSFOR you can readily appreciate our enthusiasm for successful and timely development of new, modern weapons systems. These new systems typically promise dramatic gains in combat effectiveness, and very often have the corollary advantage of reducing the troop strength needed in the operational units. Such reductions always have a cascading effect and lead to corresponding reductions in the support forces. You can look at these manpower savings in two ways. One is that reduction in required troop strength as a consequence of new modernization equipments is an ideal way of accommodating manpower cuts which Congress may be planning to impose in any event. The other is to think in terms of the dollar benefits. Personnel costs have risen drastically in recent years. For every man we can truly replace by means of better equipment, the army receives a monetary dividend of about $15,000 a year which can be applied elsewhere. Either way--the prospects of better equipment manned by fewer people are immensely attractive to the army staff and especially to the ACSFOR.

All of which brings to a key point of concern to the ACSFOR-- one of his chief problems, as I indicated earlier, is that we just aren't doing a good enough job of completing and delivering new systems on schedule. In the past few years there have been more program slippages, technical difficulties, cost growth and procurement reductions (not to mention cancellations) than anyone cares to remember. Let me show you what I mean. Recently, we reviewed the histories of a number of our major weapon systems which we are now developing and acquiring. In 1967 the Army had prepared development cost and schedule estimates for most of these and we have aggregated the projected spending plans--development costs and total planned procurement costs--for six of these systems. This is a year-by-year plot of the spending plans as of 1967. I might point out that every one of these six systems had been receiving development funding for three to five years prior to these planning estimates, and had been reviewed in detailed concept formulation and parametric design studies--so, these estimates should have been good. For the purpose of keeping it unclassified the dollar values are not shown. We have a name for this mountain of spending; we call it the "Bow-Wave". Now this Bow-Wave is very frightening to a lot of people because they see these huge dollar outlays coming and they know we can't possibly budget that kind of money-- we simply won't get that kind of money. A view of the Bow-Wave often leads to panic. It prompts some people to say--usually the budget and finance manager--we must immediately cut back on research and development starts. On the other hand, the R&D people, though, say we can't do that because we must have a large number of starts to insure getting a reasonable number of successful programs. They suggest some other remedy, like reducing the requirement or basis of issue. Now both of these points of view are right to some extent--but they are both too extreme--and they really don't get to the heart of the matter. You see, we know by experience what happens

to this Bow-Wave, and we know that we don't need to be frightened or to panic. (Figure 4)

This figure shows what that 1967 Bow-Wave looks like today. The shaded area is now history and represents funds that were not spent. It is clear that we have one very effective corrective tool in the PPBS System--it won't let us spend more than we get. But the fact that we will be forced to live within that constraint isn't much comfort because we know there is another phenomenon working therein. We know that in addition to the control which we can exercise over spending, that much of the receding of the wave is caused by program slippages. So we know we don't need to be frightened by the Bow-Wave, but we also know we must learn its anatomy and understand it better; unless we do we're certain to receive bad news and very difficult problems at fairly regular intervals. You can imagine the kinds of havoc to our attempts for the orderly planning of future force development. If we can develop more realistic curves of this nature, and understand them better we will be able to plan better.

Now let's look at that original Bow-Wave and consider one of the main reasons why it was so large in the first place, and we are convinced that risk, and the lack of risk analysis, is one of the principal reasons. Invariably when we start these programs we typically produce a single schedule with successive milestones based on achieving our objectives at each step of the way. We tend to project complete-success programs (and there is pressure to do more of the same). If, instead, we were able to effectively consider the risk at each step of the way we could produce schedules that reflect the kinds of uncertainty we always encounter. Rather than a schedule that states unequivocally that we will field a system in X years, we would much rather know the number of years until we can be, say, 90 percent confident that the system will be ready for fielding. With good estimates of this kind the ACSFOR could live much more comfortably with the Bow-Wave. With improved estimates of the risks involved he would probably not see a Bow-Wave as ominous as the one shown; and he would be in a much better position to plan the management of those trade-offs that the budgetary constraints impose. In other words, we would much prefer to program deliberate slippages, when necessary, than having them imposed on us in an unexpected and capricious manner.

So this is the challenge which has faced the ACSFOR and one which is passed to you. Give us better ways of estimating the risks that go with our program predictions so that we can learn to live and work better with this Bow-Wave. General Williams can then get on with restructuring the army in accordance with the anticipated introduction of new equipments--and at the same time be fairly confident that he's not just making trouble for his successor. (Figure 5)

I'd like to touch just briefly on another reason, separate from the impact on force planning, why it is so important to improve on our weapons acquisition process. In the years ahead we can expect to have only 3 to 4 percent of the army budget, to accomplish our equipment modernization objectives. Considering how costly new systems have become and the austere level of this modernization funding, it is apparent that we have

1967 SPENDING ESTIMATES
DEVELOPMENT AND PROCUREMENT OF SIX MAJOR WEAPON SYSTEMS

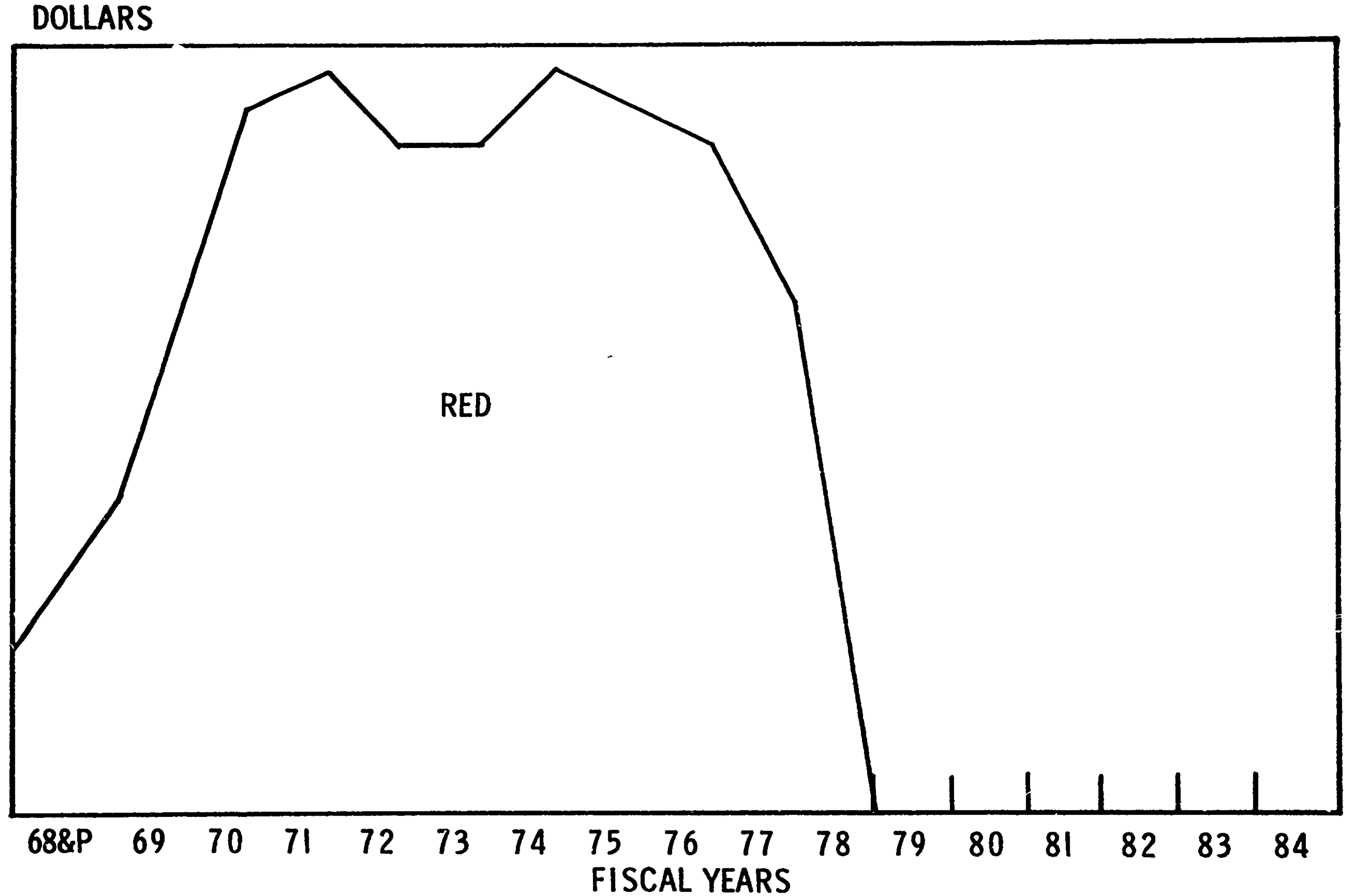


FIGURE 4.

COMPARISON OF 1967 AND 1972 COST ESTIMATES
FOR
DEVELOPMENT AND ACQUISITION OF SIX MAJOR WEAPON SYSTEMS

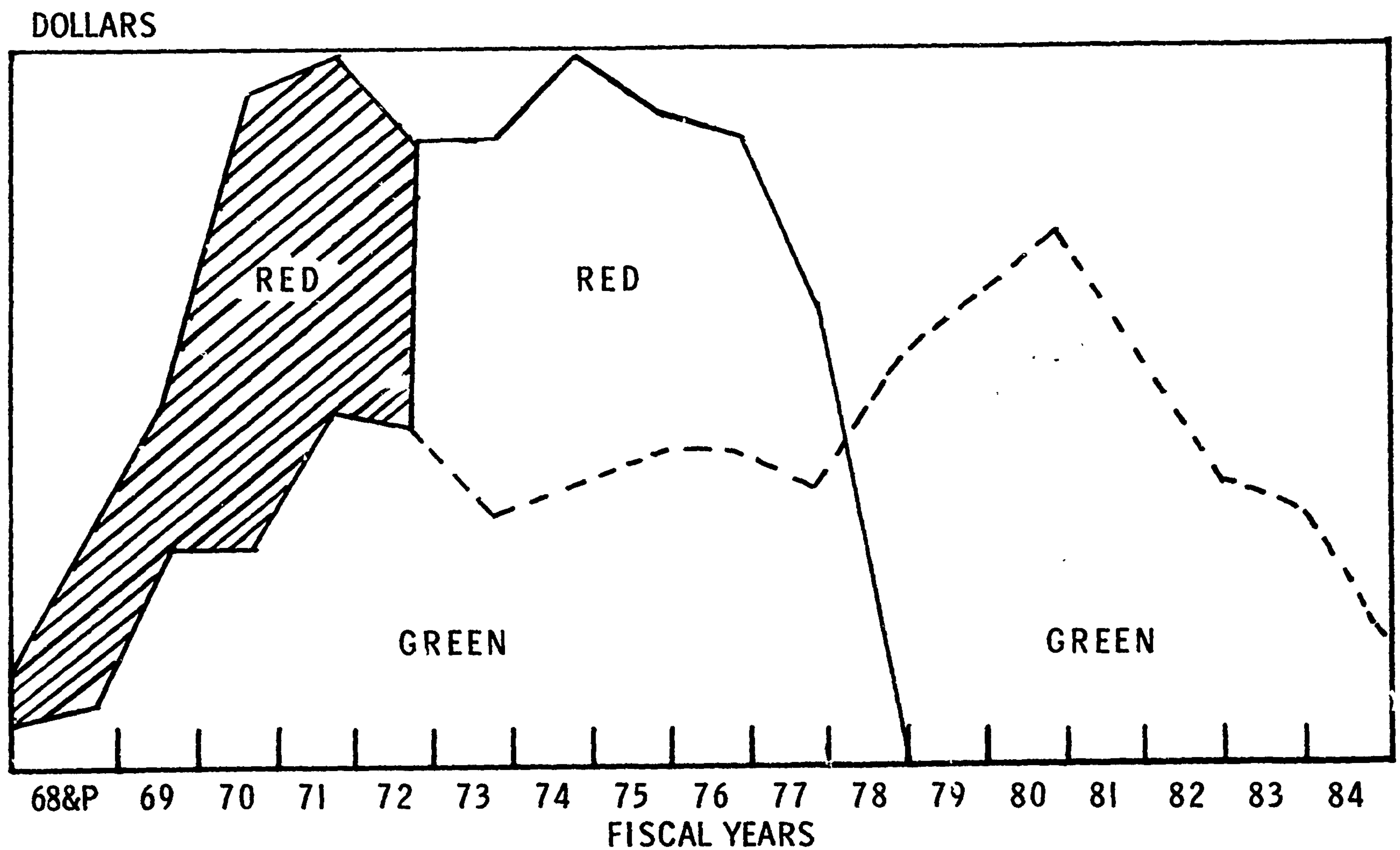


FIGURE 5.

no latitude for poorly conceived or mistaken ventures. We simply will not be able to afford cancellations or major reversals of programs in mid-stream. That is reason enough for us to take careful, measured steps toward modernization: Always fully cognizant of the risk along the path, and fully prepared with alternatives to surmount or by-pass the risk that transforms into a technical set-back.

In conclusion, I hope these remarks have added still a little extra incentive to you and your colleagues to accept the challenge to advance and successfully apply the techniques of risk analysis. I am convinced that by focusing proper attention on each and every program activity and event, the likelihood of adverse surprises can be greatly reduced. Then when problems and changes do materialize, they will have been anticipated and viable alternatives (worked out in advance) will reduce the undesirable impacts on program activities. By extending this type of careful analysis to all the major modernization programs, we would hope and expect that the critical job of force planning could be performed with a degree of credibility and at levels of precision that are simply not possible under present circumstances.

One other point, the ACSFOR knows and appreciates that the development of these techniques often require a high level of technical content which is understood by a small number of professionals. As a result he is often surprised to learn that tools for solving problems are around for some time before they become known to those at high management levels. He believes that the gap must be filled. He has been trying to do something about it at his level, but he believes that you too must work at the problem. You must try to get the word out and in a way that it is understood by as many as possible.

Now just one personal view based upon my experience, we have asked you to perform good and effective risk analysis. There will be many pressures to discredit or minimize your work. There are too many vested interests involved who shudder at the thought that their system or development may be characterized as being too risky. You must stand up to those pressures--you must maintain you objectivity and above all maintain your cool. If you don't you will become ineffective.

Finally, General Williams recognizes that it is in the nature of this risk analysis work, as well as the other analysis, that you may not see the beneficial consequences of your work for several years. And even then you may not receive full and proper credit for there will probably be some who will say things would have gone well anyway. He asks you not to let that deter you--and wishes to assure you that many others appreciate and require your efforts.

Please accept General Williams and ny own sincere good wishes for a productive and successful symposium.

Thank you, all --

DECISION ANALYSIS

STANFORD RESEARCH INSTITUTE STAFF

During the three days of the symposium, Dr. Carl Spetzler and Mr. Ramon Zamora of Stanford Research Institute presented a detailed seminar on "Decision Analysis".

The following four articles contain the substance of that material and are reporduced herein to provide a summary of this seminar:

1. "The Foundations of Decision Analysis",* Ronald A. Howard, Stanford University.

2. "A Tutorial Introduction to Decision Theory",* D. Warner North, Stanford Research Institute.

3. "Decision Analysis: Applied Decision Theory", Ronald A. Howard, Stanford University.

4. "Decision Analysis Practice: Examples and Insights", James E. Matheson, Stanford Research Institute.

*Reprinted by permission of the Institute of Electrical and Electronic Engineers, Inc.

The Foundations of Decision Analysis

RONALD A. HOWARD, SENIOR MEMBER, IEEE

Abstract—Decision analysis has emerged from theory to practice to form a discipline for balancing the many factors that bear upon a decision. Unusual features of the discipline are the treatment of uncertainty through subjective probability and of attitude toward risk through utility theory. Capturing the structure of problem relationships occupies a central position; the process can be visualized in a graphical problem space. These features are combined with other preference measures to produce a useful conceptual model for analyzing decisions, the decision analysis cycle. In its three phases—deterministic, probabilistic, and informational—the cycle progressively determines the importance of variables in deterministic, probabilistic, and economic environments. The ability to assign an economic value to the complete or partial elimination of uncertainty through experimentation is a particularly important characteristic. Recent applications in business and government indicate that the increased logical scope afforded by decision analysis offers new opportunities for rationality to those who wish it.

Introduction

Decision analysis is a term that describes a combination of philosophy, methodology, practice, and application useful in the formal introduction of logic and preferences to the decisions of the world. There was a time less than a decade ago when suggesting that decision theory had practical application evoked only doubtful comment from decision makers. The past five years have shown not only that decision theory has important practical application, but also that it can form the basis for a new professional discipline, the discipline of decision analysis. Many of the professional aspects of the field have already been described in the literature (see Howard [1]). Here we shall concentrate on the rationale and methodology of decision analysis.

In discussing the rationale and philosophy of decision analysis, we shall focus on those concepts that are most unfamiliar to the intuitive decision maker. These concepts are generally concerned with the measurement of uncertainty and with the decision maker's reaction to it. In providing a methodology for decision analysis, we shall be concerned primarily with developing a procedural form that will be broad enough to cover the important areas of application.

The Rationale of Decision Analysis

The problem of the decision maker is to select a course of action in a world that is perceived as uncertain, complex, and dynamic. To follow a course of action is to make an irreversible allocation of resources, an act that we call making a decision. Perhaps the resource whose allocation is least reversible is time, but other resources may vie for this characteristic.

Although the development of a theory of decision that comprises uncertainty, complexity, and dynamic effects is a formidable task, such a theory would not be complete, for it often turns out that what is most perplexing to the decision maker is not the mystery of his environment, but rather the specification of his own preferences. Thus we shall discuss the rationale of decision analysis by commenting on the three topics of uncertainty, structure, and preference.

Our primary interest in the topic of uncertainty is the philosophical basis for the treatment of uncertainty according to the mathematical laws of probability. The topic structure includes the complex and dynamic interactions that may exist among the many facets of a decision problem. Finally, we shall discuss under preference not only the difficulty of assigning values, but also the necessity for a value language that will be useful in a dynamic and uncertain environment.

Uncertainty

The problem of describing uncertainty has tormented philosophers for centuries. Pascal and Fermat laid the mathematical foundations of probability over three hundred years ago, and its development continues today. It might seem obvious that this theory would be the natural medium for thinking about uncertainty. However, the obvious was not proved until the present century, when it was shown that reasonable axioms for a theory of uncertainty led directly to the mathematical theory of probability.

Subjective Probability: While virtually everyone agrees on the proper use of the probability calculus, there is considerable disagreement on the interpretation of its results. Many users of probability theory consider probability to be a physical characteristic of an object as its weight, volume, or hardness. For example, they would say that a coin "has" a probability of falling heads on any toss and that to measure this probability would merely require a large number of tosses. This view of probability is called the objective interpretation.

Another group considers probability as a measure of the state of knowledge about phenomena, rather than about the phenomena themselves. This group would say that one "assigns" a probability of heads on the next toss of a coin based on all the knowledge that he has about the coin. A coin would be "fair" if, on the basis of all available evidence, there is no reason for asserting that the coin is more

Manuscript received July 2, 1968. This research was partially supported by the National Science Foundation under Grant NSF-GK-1683 and by the Office of Naval Research under Contracts ONR N00014-67-A-0112-0008 and ONR N00014-67-A-0112-001C.

The author is with the Department of Engineering-Economic Systems, Stanford University, Stanford, Calif.

likely to fall heads than tails. This view is called the subjective interpretation.

The distinction between the interpretations might seem small, but it is the key to the power of decision analysis. The objectivist requires repeatability of phenomena under essentially unchanged situations to make what he would consider to be meaningful inferences. The subjectivist can accept any amount of data, including none, and still apply logic to the decision. The objectivist was able to survive and even flourish, when the main problems of inference arose in areas such as agriculture that provide large amounts of cheap data. Today, when decisions regarding space programs must be based on a single launch of a one hundred million dollar rocket, the ability of the subjectivist to apply logic to one-of-a-kind situations has become indispensable.

These examples might lead one to believe that the subjective view of probability is modern; in fact, it was clearly held and understood by Bayes and Laplace two hundred years ago. The objectivist view is associated primarily with the founding of the British school of statistics in the early 1900's. It is the feeling of many, including decision analysts, that the creation of the field of statistics through the advent of the objective interpretation was a heresy in the development of the treatment of uncertainty. While objectivists are definitely in the majority at present, their ranks seem to be diminishing.

Subjective Probability Notation: Since the decision analyst necessarily holds the subjective viewpoint, he prefers a notation for probability that reveals that it is an assignment based on a certain set of information. Such a notation is constructed as follows: Let A be an event and $\mathcal{S}$ be the state of information on which the probability of the occurrence of A is to be assigned. Then $\{A|\mathcal{S}\}$ is the symbol for the probability of A given $\mathcal{S}$. If x is a random variable, then the probability density or mass function of x assigned on the basis of $\mathcal{S}$ is $\{x|\mathcal{S}\}$. The expectation of x based on $\mathcal{S}$ is written $\langle x|\mathcal{S}\rangle$ and is defined by

$$\langle x|\mathcal{S}\rangle = \int_x x\{x|\mathcal{S}\}$$

where $\int_x$ is a general summation on x to be interpreted as a summation or integration as appropriate. The nth moment of x based on $\mathcal{S}$ would then be

$$\langle x^n|\mathcal{S}\rangle = \int_x x^n\{x|\mathcal{S}\}.$$

The variance of x is written ${}^v\langle x|\mathcal{S}\rangle$ and defined by

$${}^v\langle x|\mathcal{S}\rangle = \langle x^2|\mathcal{S}\rangle - \langle x|\mathcal{S}\rangle^2.$$

One very special state of information is the total knowledge available at the beginning of the problem under consideration, the total prior experience denoted by $\mathcal{E}$. Then $\{x|\mathcal{E}\}$ would be called the prior density function on x, or the "prior" for short. The quantities $\langle x|\mathcal{E}\rangle$ and ${}^v\langle x|\mathcal{E}\rangle$ would then be the prior mean and variance.

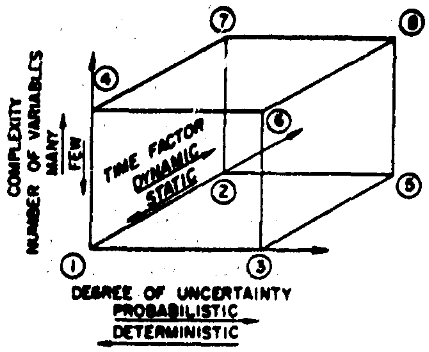


Fig. 1. Problem space.

Although this notation often seems strange, it provides a mathematical language for uncertainty that describes precisely both the quantities on which the probability assignment is to be made and the state of information to be used in the assignment. The subjective view thus induces not only care in the interpretation of probability but also precision in its written expression.

Structure

The primary function of the decision analyst is to capture the relationships among the many variables in a decision problem, a process called structuring. The complexity of structure required will differ from problem to problem: from a "back-of-the-envelope" decision tree to a system of interconnected programs that tax the largest computers.

The Problem Space: A diagram like Fig. 1 is an aid in visualization. This diagram, the problem space, permits characterizing decision problems by their underlying structure. The dimensions of the problem space are degrees of uncertainty, time dependence, and complexity. Degree of uncertainty can range from the deterministic situations, where all variables are known, to the highly probabilistic situations, where little information is available about any problem variables. The time dependence can range from static to dynamic; complexity is measured in terms of the number of variables required.

Each corner of the problem space corresponds to certain mathematical models. Corner 1 is the deterministic static one-variable decision problem, such as that of finding the largest rectangular area that can be fenced with a given length of fencing. The models of elementary calculus, developed over 300 years ago, would be appropriate. Corner 2, the deterministic dynamic single-variable decision problem, would arise in elementary automatic control applications. The mathematical models of differential equations and transform calculus would be relevant; they were developed over 100 years ago. Corner 3 represents the probabilistic static single-variable problem, such as whether or not to buy life insurance. Three-hundred-year-old elementary probability would be quite helpful in reaching a decision. Corner 4 introduces complexity in the form of the deterministic static, but many-variable problem. Decision problems like assigning customers to warehouses or men to jobs provide an illustration.

One-hundred-year-old matrix algebra and 20-year-old linear optimization techniques would be very useful.

Corner 5 combines the two factors of uncertainty and dynamism in the uncertain dynamic, but single-variable problem, such as simple inventory control. Here the theory of stochastic processes and queuing models developed over the last 50 years would be most relevant. Corner 6 corresponds to the probabilistic static multivariable problem. Decision problems like bidding on new product introduction might have such an underlying structure. The mathematics of joint probability distributions would be especially helpful. Corner 7 refers to the deterministic dynamic multivariable decision problem, such as the complicated control problems posed by a space vehicle or a steel mill. Although probabilistic elements may be present, they are usually treated as perturbations of the deterministic model. The modern theory of control developed in the past three decades applies successfully to these problems.

Finally, corner 8 is the most complex corner, describing problems involving uncertainty, dynamism, and complexity. In a sense, all decision problems could be located here because they all involve the three factors to some degree. However, this corner is used to indicate problems where the three elements are indispensable to a meaningful analysis. Problems like electrical power system planning or business mergers are particular examples. Useful models might be Markov processes and their derivatives.

The extent to which formal models are available varies considerably over the problem space. Near the origin there are usually several alternative models for the problem; near corner 8 it is more a matter of patching together approximations to obtain a useful representation. As technology advances, more realistic models of uncertain, dynamic, and complex processes will be developed. However, it will continue to be the job of the decision analyst to be the engineer who matches technology to the requirements of the problem. His product is the embodiment of logic.

Preference

The problem of preference measurement is to determine in quantitative terms just what the decision maker wants.

Value: The first step is to assign a single value v to each possible outcome of the decision problem. If the problem is concerned with the allocation of monetary resources, then it is logical to measure this value in monetary terms. In business organizations, some form of profit may be appropriate. But the need for monetary values as a precedent for monetary allocation applies even if the outcome involves the loss of life or limb. As decision analysis is increasingly used in problems of social significance, a monetary value may have to be assigned to such outcomes as a cultured life or an ignorant life. Though these assignments may be very difficult, there is no rational alternative.

Time Preference: However, even in dynamic world, the preference question would not be resolved until the decision maker had stated his preference for outcomes that are distributed in time: a preference called time preference. The importance of time preference is revealed when the analyst studies problems like the development of the national parklands or management of an individual's investment portfolio.

The phenomenon of time preference could be described as the greed–impatience tradeoff. It is characteristic of individuals and organizations that they want more now. However, the alternatives provided often give them a choice between more later or less now. Examples would be the choice between hydroelectric and gas turbine electricity production or, in general, the choice between investment in capital goods and consumer goods.

While the problem of preference is complicated, it is usually treated in decision analysis by the specification of a discount or interest rate and the rule that the alternative with the highest discounted, or present, value is to be preferred. Even within this framework, selecting the appropriate interest rate is not easy; it involves the nature of the interaction between the organization and its financial environment.

Risk Preference: The most unusual and challenging preference problem concerns preference for risk. The existence of the phenomenon is established by noting that few people are willing to bet double or nothing on next year's salary, even though the proposition is fair. Most people and organizations are averse to risk: they are willing to engage in uncertain propositions only if the expected value of the proposition is positive and relatively large. The description of this type of preference requires a set of concepts that are unusual, but logical.

To be specific in describing the concepts, it is necessary to define the technical term "lottery." A lottery is a set of prizes or prospects, one and only one of which will be received. Associated with each prize is a probability; the sum of all the probabilities is one. In many cases the prizes will each correspond to the amount of some commodity, such as money, that will be received. In these cases we can think of the lottery as a random variable described by either a probability mass or probability density function.

Utility theory: The most common structure for encoding risk preference requires that the individual subscribe to a set of axioms concerning lotteries. The first is that he must be willing to provide a transitive rank ordering of all prizes in any lottery. That is, if the prizes in a lottery are A, B, and C, he must be able to say in what order he prefers the prizes; further, if he prefers A to B and B to C, then he must prefer A to C.

The second axiom is that if he says he prefers A to B to C then there must exist a value of p such that he is indifferent between receiving B for certain and participating in a lottery that produces A with probability p and C with probability $1 - p$. When the appropriate value of p has been found, we would say that B is the certain equivalent of the lottery on A and C.

The third axiom is that if he prefers prize A to prize B and if he is presented with two lotteries, each offering A

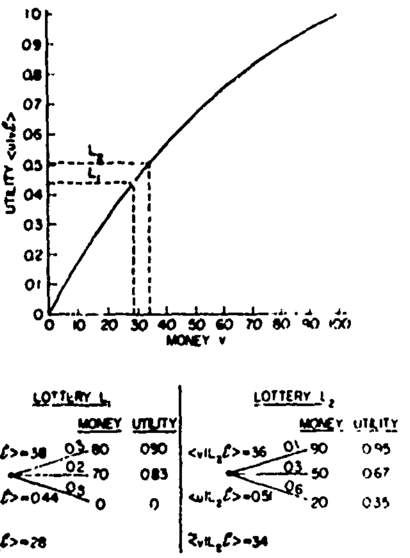


Fig. 2. Utility curve.

and B with different probabilities, then he must prefer the lottery that yields A with the higher probability.

These axioms are the most significant ones. However, two others are necessary for completeness. One is that a certain equivalent of a lottery may be substituted for the lottery in any situation without changing the preferences of the decision maker; we might call this a "did you really mean it?" axiom. The other is that a lottery whose prizes are themselves lotteries is equivalent to a lottery that produces the same ultimate prizes with probability computed according to the laws of probability; this could be termed a "no fun in gambling" axiom.

Mathematical arguments reveal that an individual who subscribes to these axioms can encode his risk preference in terms of a function on the prizes of the lotteries, a function called a utility function. The utility function has two important properties: first, that the utility of any lottery is the expected utility of its prizes; second, that if one lottery is preferred to another by the individual, then its utility will be higher.

Thus the utility function assigns to any lottery a real number; the lotteries will be preferred in the order of these numbers. However, the actual magnitude of the utility is not important, because the preferences revealed by the utility function are unchanged if the utility function is modified by multiplication by a positive constant or by addition of any constant. Thus the utility function serves as a risk preference thermometer that can be used for ranking lotteries according to the risk preference of an individual.

In problems of professional interest the lottery prizes are usually measured in a commodity such as money. In this case the utility function can be represented by a curve that shows the utility to be assigned to any amount of the commodity. Such a utility curve appears as Fig. 2. The curve $\langle u|v\mathcal{E}\rangle$ shows the utility u assigned by some individual to amounts of money v between 0 and 100 dollars. Because of the invariance to linear transformation, the scale of measurement can be selected arbitrarily; this curve assigns a utility of 0 to 0 dollars and a utility of 1 to 100 dollars.

The two lotteries below the curve show how it is used. The expected value of a lottery L is defined in our notation by

$$\langle v|L\mathcal{E}\rangle = \int_v v\{v|L\mathcal{E}\}.$$

Lottery L_1 has an expected value of 38 dollars; L_2, an expected value of 36 dollars. Someone who was indifferent to risk would prefer L_1. However, to determine the preference of the individual with the utility function in Fig. 2, we first determine the utility of each prize in each lottery from the utility curve and then find the expected value of the utility. The expected utility of a lottery is given by

$$\langle u|L\mathcal{E}\rangle = \int_v \langle u|v\mathcal{E}\rangle\{v|L\mathcal{E}\}.$$

Since the expected utility of lottery L_1 is 0.44, while that of lottery L_2 is 0.51, the individual would prefer lottery L_2, in spite of its lower expected value. We would describe individuals whose utility curves are concave downwards as risk averse.

The certain equivalent: Although this calculation serves to determine the individual's preference, it gives us no feeling about the strength of the preference. The magnitude of the utility can be no help because we see that if we added 10 to all utility numbers, we would derive exactly the same preference ordering but with much smaller percentage difference in utility numbers. To measure strength of preference, it is helpful to return to the concept of certain equivalent.

To evaluate a lottery in a single meaningful monetary number, we can ask what amount of money received for certain would have the same utility as the lottery. The certain equivalent of a lottery L, denoted by $\sim\langle v|L\mathcal{E}\rangle$, is thus the amount of money shown by the utility curve to have the same utility as the lottery. The certain equivalent is mathematically defined by the equation

$$\langle u|v = \sim\langle v|L\mathcal{E}\rangle, \mathcal{E}\rangle = \langle u|L\mathcal{E}\rangle.$$

Thus from the curve we see that the utility of 0.44 for lottery L_1 corresponds to a certain equivalent of 28 dollars, while the utility of 0.51 for lottery L_2 would mean a certain equivalent of 34 dollars. The individual would be just indifferent between receiving either 28 dollars for certain or lottery L_1 and between receiving 34 dollars for certain or lottery L_2. It would be slightly inaccurate, but intuitively satisfying, to say that lottery L_2 is worth 6 dollars more to the individual than is lottery L_1.

Exponential utility curves: In some cases the individual is willing to subscribe to a sixth axiom: that if all prizes in a lottery are increased by any amount Δ, the certain equivalent of the lottery will also increase by Δ. The axiom is persuasive, since the increment Δ will be received with certainty regardless of the outcome of the lottery. However,

the axiom is very powerful, for someone who subscribes to it must have a utility curve that is linear or exponential in form; that is, $\langle u|v\mathcal{E}\rangle$ is proportional either to v or to $e^{-\gamma v}$. Furthermore, the curve is completely described by the constant γ called the risk aversion coefficient. Although few individuals may in fact wish to be governed by this axiom, the exponential utility curve is very useful in analyses, as we shall see.

Stochastic dominance: There is one important case in which risk preference need not be measured at all. That is the case in which the choice between two alternatives would be clear to a rational man regardless of his risk preference; it is called the case of stochastic dominance. Lottery L_1 stochastically dominates lottery L_2 if the probability of receiving a monetary return in excess of c is higher for L_1 than for L_2 for any value of c; that is,

$$\{v > c|L_1\} > \{v > c|L_2\}, \quad -\infty < c < \infty.$$

If one lottery stochastically dominates all others, then it will be preferred by the individual regardless of his attitude toward risk; there is no need to use the utility function.

Joint Time–Risk Preference: Individuals often have to choose between monetary rewards that are not only uncertain, but distributed over time. In these situations time and risk preference must be jointly encoded. The description of joint time–risk preference is a problem that admits many solutions. Here we shall employ the idea of reducing any time stream of value to a present value using the time preference measure and then applying the utility function to determine which lottery on present values is most desirable.

The Methodology of Decision Analysis

With this background we can go on to a discussion of how a decision problem can be progressively analyzed using decision analysis principles. The procedure is best explained in terms of a diagram like that in Fig. 3. Here we view the decision analysis procedure as divided into three major phases, the deterministic, probabilistic, and informational phases. The deterministic phase establishes the deterministic relationships among the variables of the problem. The probabilistic phase introduces uncertainty and risk preference. Finally, the informational phase determines the economic value of gathering more information. Following these phases, a decision is required on whether to act or to gather new information. If additional information is obtained, e.g., through market testing or building a pilot plant, then this information must be incorporated into the structure and probability assignments of the problem; the cycle is then repeated.

The decision analysis cycle is a convenient conceptual model rather than an inevitable method for analyzing decision problems. With this point in mind, we shall now examine the steps required in each phase.

The Deterministic Phase

The first step in the deterministic phase is to construct a deterministic model of the decision problem.

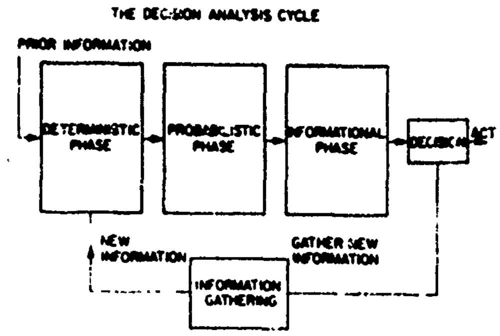


Fig. 3. Decision analysis cycle.

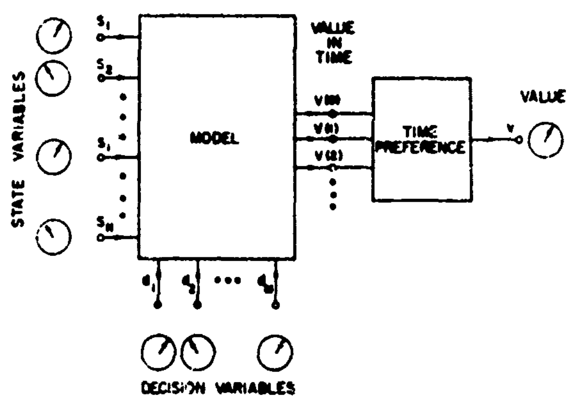


Fig. 4. Deterministic model.

The Deterministic Model: Fig. 4 is an abstract representation of the model. The model relates the important variables in the problem that are not under the control of the decision maker and the variables that are under his control to the production of value in time. These variables are called the state variables s_i and decision variables d_i. We can visualize the state variables as a set of knobs on the model that are set by a disinterested nature; the decision variables are knobs set by the decision maker. Fig. 4 shows that the values developed over time $v(0)$, $v(1)$, $v(2)$, $\cdots$ are operated upon by the time preference specification to produce a present value reading v that we may regard as appearing on a value meter. Thus any setting of the state- and decision-variable knobs will produce a value reading. The deterministic model will generally be realized in the form of a computer program.

Deterministic Sensitivity: Fig. 5 shows the first analytical step in the deterministic phase, the measurement of deterministic sensitivity. In the representation of Fig. 5 the time preference measure is shown incorporated into the deterministic model to produce a single present value reading. The analysis begins by assigning each state variable a nominal value and a range that might correspond to the 10- and 90-percent point on its marginal cumulative probability distribution. Decision variables would also be assigned nominal values and ranges to reflect initial feelings about what the best decision might be.

With all variables but one set to their nominal values, that one variable would be swept across its range to deter-

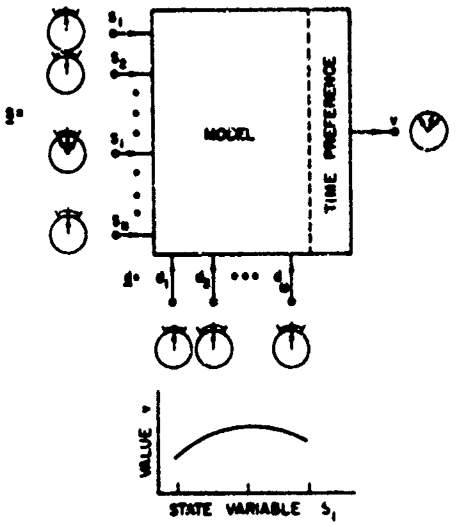


Fig. 5. Deterministic sensitivity.

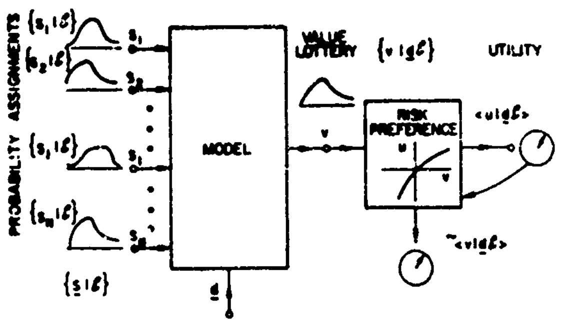


Fig. 6. Value lottery.

mine the effect on the value reading. The figure shows the measurement for the ith state variable s_i. State or decision variables that showed high sensitivity would be retained in the further analyses of the model. A variable could show a high deterministic sensitivity because of its wide range, crucial nature, or a combination of these effects.

In some problems this one-at-a-time type of sensitivity analysis will not be sufficient: the joint sensitivity of variables will have to be measured by sweeping more than one variable at a time over their ranges. Because the number of possibilities for joint sensitivity increases combinatorially with the number of variables, the analyst must use judgment in determining where joint sensitivity measurements will be required.

The net effect of the deterministic sensitivity analysis will be to determine the state variables and decision variables that have a major effect on value. The next step will be to introduce the current state of knowledge on uncertainty in the state variables and determine which decision would be best, given the uncertainty; this is done in the probabilistic phase.

The Probabilistic Phase

The probabilistic phase requires assignment of probability distributions on the state variables.

The Value Lottery: Fig. 6 shows this assignment as a marginal probability distribution $\{s_i|\mathcal{E}\}$ on each state variable. Since the state variables will generally be jointly related, the complete description of the state of knowledge about them would be the joint probability distribution $\{s_1,s_2,\cdots,s_N|\mathcal{E}\} = \{\mathbf{s}|\mathcal{E}\}$, but the marginal distributions shown will serve as a pictorial representation. The settings of the decision variables are summarized by the decision vector $\mathbf{d} = [d_1,d_2,\cdots,d_M]$. For any setting $\mathbf{d}$ the joint distribution $\{\mathbf{s}|\mathcal{E}\}$ on the state variables will imply a probability distribution on the value, $\{v|\mathbf{d}\mathcal{E}\}$, a distribution we call the value lottery. The decision problem then reduces to finding the setting $\mathbf{d}$ that produces the most desirable value lottery.

The determination of the value lottery corresponding to any decision vector $\mathbf{d}$ will be performed by analytical or simulation methods, as appropriate. Efficient search procedures are helpful in establishing the best setting for $\mathbf{d}$.

Risk Preference: There remains the question of which value lottery is best. Perhaps the question will be easily resolved by the observation that one setting of $\mathbf{d}$ produces a value lottery that stochastically dominates the lotteries produced by all other settings. But if not, then it will be necessary to encode the risk preference of the decision maker in a utility curve. This curve will allow each value lottery and hence each setting of $\mathbf{d}$ to be rated by its utility. The setting that produces the highest utility $\langle u|\mathbf{d}\mathcal{E}\rangle$ would then be judged the best. To gain intuitive meaning, the utility of each lottery could be returned to the utility curve to show the certain equivalent value $\sim\langle v|\mathbf{d}\mathcal{E}\rangle$ implied by the decision setting $\mathbf{d}$.

This procedure establishes the setting of the decision variables $\mathbf{d}(\mathcal{E})$, that is most desirable to the decision maker in view of his state of knowledge regarding uncertainties and his risk preferences,

$$\mathbf{d}(\mathcal{E}) = \max_{\mathbf{d}}{}^{-1}\langle u|\mathbf{d}\mathcal{E}\rangle = \max_{\mathbf{d}}{}^{-1} \sim\langle v|\mathbf{d}\mathcal{E}\rangle.$$

Furthermore, it shows the utility $\langle u|\mathcal{E}\rangle$ and certain equivalent $\sim\langle v|\mathcal{E}\rangle$ of the best decision,

$$\langle u|\mathcal{E}\rangle = \langle u|\mathbf{d} = \mathbf{d}(\mathcal{E})\mathcal{E}\rangle$$

$$\sim\langle v|\mathcal{E}\rangle = \sim\langle v|\mathbf{d} = \mathbf{d}(\mathcal{E})\mathcal{E}\rangle.$$

In a sense, this step completes the solution of the decision problem. However, since decision analysis is more engineering than mathematics, the procedure does not stop here, but rather continues to the measurement of another kind of sensitivity, stochastic sensitivity.

Stochastic Sensitivity: The idea behind stochastic sensitivity is the desire to measure the effect of a variable on the result of the decision problem not in the deterministic environment where all other variables are set to their nominal values, but in the probabilistic environment where all other variables are governed by their appropriate probability distributions. As Fig. 7 shows, if the ith state variable s_i were known, the other state variables would be governed by the conditional distribution $\{\mathbf{s}|s_i\mathcal{E}\}$ obtained by dividing $\{\mathbf{s}|\mathcal{E}\}$ by $\{s_i|\mathcal{E}\}$. Thus the specification of any

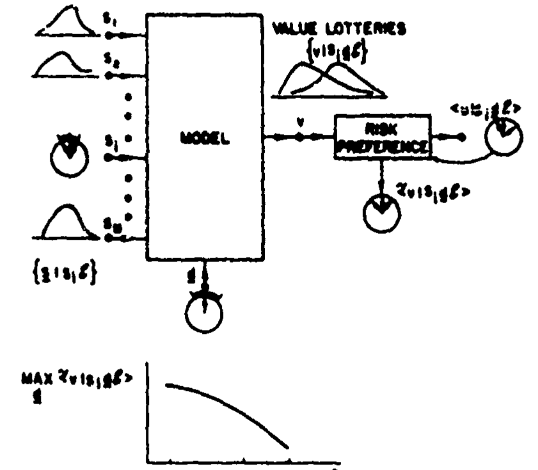


Fig. 7. Stochastic sensitivity.

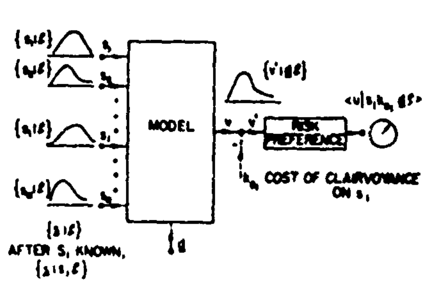


Fig. 8. Clairvoyance.

value for s_i would imply some joint probability distribution of the remaining state variables, and in turn a value lottery $\{v|s_i d\mathcal{E}\}$ for the given setting of the decision vector. The risk preference encoding would describe this value lottery by a certain equivalent $\sim\langle v|s_i d\mathcal{E}\rangle$.

Suppose now that the decision vector $\boldsymbol{d}$ is adjusted to the value $\boldsymbol{d}(s_i\mathcal{E})$ that produces the highest certain equivalent for this value of s_i, $\max_{\boldsymbol{d}} \sim\langle v|s_i d\mathcal{E}\rangle$, that is,

$$\boldsymbol{d}(s_i\mathcal{E}) = \max_{\boldsymbol{d}}{}^{-1}\langle u|s_i d\mathcal{E}\rangle = \max_{\boldsymbol{d}}{}^{-1} \sim\langle v|s_i d\mathcal{E}\rangle.$$

If this procedure is repeated for the various values of s_i within its range, the plot of $\max_{\boldsymbol{d}} \sim\langle v|s_i d\mathcal{E}\rangle$ will show the stochastic sensitivity of the variable s_i.

Stochastic sensitivity shows how the certain equivalent of the decision problem depends on a particular state variable when all other state variables are uncertain. Stochastic sensitivity can be measured in a different sense if, rather than choosing the best decision variable setting $\boldsymbol{d}$ for each s_i, the setting $\boldsymbol{d}(\mathcal{E})$ that was best for $\{\boldsymbol{s}|\mathcal{E}\}$ is used throughout. This technique measures the stochastic sensitivity to the ith state variables under the original decision rule rather than under a decision rule adjusted to take advantage of knowledge of s_i. Stochastic sensitivity to a decision variable d_i can be measured by using the probability assignment $\{\boldsymbol{s}|\mathcal{E}\}$ for the state variables and then seeing how the certain equivalent changes with d_i either with other decision variables fixed or continually optimized.

The problems of joint sensitivity measurement arise just as they did in the case of deterministic sensitivity. However, here the cost of joint sensitivity measurement is even greater than before because of the need to develop lotteries on value rather than single numbers.

Stochastic sensitivity can provide important additional insight into problem relationships. It can show the need for further structure to allow available information to be encoded more effectively. It might reveal that variables originally thought to be of vital importance on the basis of deterministic analyses are relatively unimportant in the probabilistic environment. At a minimum, it yields a useful measurement of the robustness of the indicated decision.

The Informational Phase

The probabilistic phase of the analysis provides further insight into the importance of uncertainty in state variables, but it stops short of what we would really like to know, namely, what is the worth in monetary terms of the various forms of uncertainty remaining in the problem? The informational phase covers this last step of measuring economic sensitivity and hence indicates what sort of additional information could be economically gathered.

Clairvoyance: A useful concept in discussing the informational phase will be the clairvoyant. The clairvoyant is an individual who can tell us the precise value of any uncertain variable. Clearly, such help would be valuable, but how valuable?

Fig. 8 illustrates the case where we have engaged the clairvoyant to tell us the value of the ith state variable s_i at a cost k_{s_i}. Knowing s_i will have two effects on the result. First, the probability assignments on the other state variables will be governed by $\{\boldsymbol{s}|s_i\mathcal{E}\}$. Second, whatever present value v is produced will have to be reduced by the clairvoyant's charge k_{s_i} to a net present value v'. Once s_i is reported, the best setting $\boldsymbol{d}(s_i k_{s_i}\mathcal{E})$ of the decision vector will be the setting that produces a net present value lottery having the highest utility. Thus

$$\boldsymbol{d}(s_i k_{s_i}\mathcal{E}) = \max_{\boldsymbol{d}}{}^{-1}\langle u|s_i k_{s_i} d\mathcal{E}\rangle = \max_{\boldsymbol{d}}{}^{-1} \int_{\boldsymbol{s}} \langle u|\boldsymbol{s} k_{s_i} d\mathcal{E}\rangle \{\boldsymbol{s}|s_i\mathcal{E}\}$$

and

$$\max_{\boldsymbol{d}} \langle u|s_i k_{s_i} d\mathcal{E}\rangle = \langle u|s_i k_{s_i} \boldsymbol{d}(s_i k_{s_i}\mathcal{E})\mathcal{E}\rangle.$$

Therefore, if we knew that the clairvoyant would report a particular value of s_i, the utility of the resulting lottery would be $\langle u|s_i k_{s_i} \boldsymbol{d}(s_i k_{s_i}\mathcal{E})\mathcal{E}\rangle$. However, we are not sure that he will report that value; indeed, if we were sure, there would be no point in employing him. Consequently, we must weight the utility we shall derive if he reports a value of s_i by the probability that he will report that value in order to determine the utility $\langle u|k_{s_i}\mathcal{E}\rangle$ of the lottery we enter by engaging him. The probability we assign to his reporting any value of s_i is, of course, just $\{s_i|\mathcal{E}\}$ since he is assumed competent and trustworthy. Therefore,

$$\langle u|k_{s_i}\mathcal{E}\rangle = \int_{s_i} \langle u|s_i k_{s_i} \boldsymbol{d}(s_i k_{s_i}\mathcal{E})\mathcal{E}\rangle \{s_i|\mathcal{E}\}.$$

If the cost of the clairvoyant k_{s_i} were equal to zero, we would expect this utility $\langle u|k_{s_i} = 0\,\mathcal{E}\rangle$ to be greater than the utility $\langle u|\mathcal{E}\rangle$ of the best lottery without clairvoyance. However, as the cost of the clairvoyant increases, his service will become progressively less desirable until the utility of the lottery with clairvoyance is just equal to the utility of the best lottery without clairvoyance. The value of k_{s_i} that satisfies the equation

$$\langle u|k_{s_i}\mathcal{E}\rangle = \langle u|\mathcal{E}\rangle$$

is called the value of clairvoyance about the variable s_i.

The value of clairvoyance on a variable is an important quantity because it represents the largest amount that one should pay to eliminate completely uncertainty regarding the variable. Since most real information gathering opportunities provide less than perfect information, they should never be employed when their cost exceeds the cost of clairvoyance.

Notice that the actual availability of a clairvoyant is irrelevant to this argument. The clairvoyant in decision analysis plays exactly the same role as the Carnot engine in thermodynamics: a conceptual reference against which to compare the performance of physically realizable alternatives.

As with sensitivity measurement, the value of simultaneous clairvoyance on several variables can also be calculated with somewhat more difficulty. In the preceding argument, s_i would be replaced by a subset of state variables, but the nature of the calculations remains the same. Even if the state variables are independent, the value of clairvoyance on several of them can differ from the sum of the values of clairvoyance on each separately. (See [4], [5].)

The value of clairvoyance on any state variable or set of state variables will depend on the prior distribution $\{\mathbf{s}|\mathcal{E}\}$. It is clear that some prior distribution will maximize the value of clairvoyance; we might call this the maximum value of clairvoyance. It is the value of clairvoyance to a decision maker who had the most unfortunate initial state of information as far as purchasing clairvoyance is concerned. The calculation is useful because it shows the most that anyone should pay for clairvoyance regardless of his state of information. Of course, the calculation is predicated on a given time and risk preference.

Experimentation: The real-world approximation to clairvoyance is some form of experimentation. An important question in guiding the gathering of additional information is, therefore, the value of a given experiment. The calculation follows almost the same form as the computation of the value of clairvoyance.

Fig. 9 illustrates the nature of the calculation. Suppose that the experiment costs $k_{\mathbf{E}}$ and that after it was conducted, it produced the data D. Knowledge of D would change the probability distribution on $\mathbf{s}$ to $\{\mathbf{s}|D\mathcal{E}\}$, which is related to the prior distribution $\{\mathbf{s}|\mathcal{E}\}$ by Bayes' equation,

$$\{\mathbf{s}|D\mathcal{E}\} = \frac{\{D|\mathbf{s}\mathcal{E}\}\,\{\mathbf{s}|\mathcal{E}\}}{\{D|\mathcal{E}\}}.$$

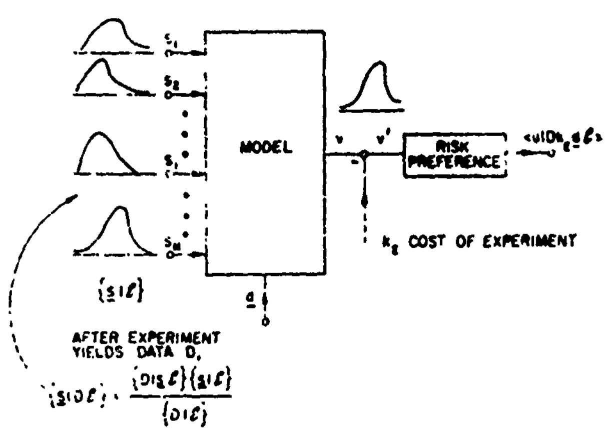


Fig. 9. Experimentation.

The new quantities $\{D|\mathbf{s}\mathcal{E}\}$ and $\{D|\mathcal{E}\}$ are interesting in themsevles. The quantity $\{D|\mathbf{s}\mathcal{E}\}$ is the probability of observing the particular data D for any setting of the state variables; it is called the likelihood function. The quantity $\{D|\mathcal{E}\}$ is the probability of observing D assigned before the experiment is performed; it is related to the likelihood function and the prior by

$$\{D|\mathcal{E}\} = \int_{\mathbf{s}} \{D|\mathbf{s}\mathcal{E}\}\,\{\mathbf{s}|\mathcal{E}\}$$

and is called the preposterior distribution.

Once D is known, the best setting $\mathbf{d}(Dk_{\mathbf{E}}\mathcal{E})$ of the decision vector will be the setting that produces the net present value lottery of highest utility,

$$\mathbf{d}(Dk_{\mathbf{E}}\mathcal{E}) = \max_{\mathbf{d}}{}^{-1}\langle u|Dk_{\mathbf{E}}\mathbf{d}\mathcal{E}\rangle$$

$$= \max_{\mathbf{d}}{}^{-1} \int_{\mathbf{s}} \langle u|\mathbf{s}k_{\mathbf{E}}\mathbf{d}\mathcal{E}\rangle\,\{\mathbf{s}|D\mathcal{E}\}.$$

The utility of this lottery will be

$$\max_{\mathbf{d}}\langle u|Dk_{\mathbf{E}}\mathbf{d}\mathcal{E}\rangle = \langle u|Dk_{\mathbf{E}}\mathbf{d}(Dk_{\mathbf{E}}\mathcal{E})\mathcal{E}\rangle.$$

However, this utility will be received conditional on the reporting of D. The probability that D will be reported by the experiment is the preposterior probability $\{D|\mathcal{E}\}$. Therefore, the overall utility of the experiment at a cost $k_{\mathbf{E}}$, $\langle u|k_{\mathbf{E}}\mathcal{E}\rangle$, will be just

$$\langle u|k_{\mathbf{E}}\mathcal{E}\rangle = \int_D \langle u|Dk_{\mathbf{E}}\mathbf{d}(Dk_{\mathbf{E}}\mathcal{E})\mathcal{E}\rangle\,\{D|\mathcal{E}\}.$$

The number $k_{\mathbf{E}}$ that satisfies the equation

$$\langle u|k_{\mathbf{E}}\mathcal{E}\rangle = \langle u|\mathcal{E}\rangle$$

and thus makes the utility of the best lottery with the experiment equal to the utility of the best lottery without the experiment is the value of the experiment.

Comparing this calculation with the one for the value of clairvoyance shows that we can interpret clairvoyance as a very special kind of experiment: one that completely eliminates uncertainty in one or several state variables.

Once the value of the experiment has been computed, it can be compared with its real-world cost. Experiments whose value exceeds their cost are profitable alternatives for the decision maker; others are not. Determining the profitability of various information gathering plans shows which, if any, should be pursued before the primary decision is made.

The Decision Analysis Cycle

This discussion of the decision analysis cycle has indicated most, but not all, of the types of analyses that may be useful. For example, determining sensitivity of the best decision and its present value to the discount rate representing time preference would be an obvious test to perform. In some decision problems, particularly those requiring the consensus of several interested parties, it may be wise to measure risk sensitivity. This would involve seeing how the best decision and its certain equivalent value change as the risk aversion coefficient is increased. Fortunately, it often happens that the same policy remains best for a range of risk coefficients that includes those of all participants. In these cases, there is no point in argument over just what attitude toward risk should govern the decision.

Division of Effort: The total effort devoted to the cycle is not typically equally divided among the phases. Because of the need for a detailed understanding of fundamental problem relationships, the deterministic phase requires about 60 percent of total effort. The probabilistic phase might receive 25 percent; the informational phase, the remaining 15 percent. As the analysis progresses through the phases, the nature of the work changes from the construction and tuning of the model to the development of insight by exercising it.

Computational Demands. The difficulty of exercising the model changes from phase to phase. For example, a computer run to establish stochastic sensitivity might require ten times as much time as a run to measure deterministic sensitivity. Similarly, an economic sensitivity run in the informational phase might require ten times as much computation as the measurement of stochastic sensitivity. Thus we see the need for the continued screening of variables to assure that only important factors are retained in each phase of the analysis. To think of performing a decision analysis by including all possibly relevant variables in each phase would be very unrealistic.

The Model Sequence: Typically, a decision analysis is performed not with one, but with a sequence of progressively more realistic models. The first model in the sequence we call the pilot model; it is an extremely simplified representation of the problem, useful only for determining the most important relationships. Its aeronautical counterpart would be the wind tunnel model of a new airplane. It looks very little like the desired final product, but it is indispensable in achieving that goal. Perhaps 20 percent of total effort might be devoted to construction and testing of the pilot model.

The next model in the sequence is called the prototype model. It is a quite detailed representation of the problem that may, however, still be lacking a few important attributes. Its aeronautical analogy would be the first flying model of a new airplane. While it will generally have bugs that must be eliminated, it does demonstrate overall appearance and performance of the final version. Because of the need for verisimilitude of the prototype model, it might require 60 percent of the total effort.

The final model in the sequence is the production model; it is as accurate a representation of reality as will be produced in the decision analysis. Like the production airplane, it should function well even though it may retain features that are treated in a less-than-ideal way. Perhaps 20 percent of the total effort might be devoted to this final stage of model development. When completed, the production model should be able to withstand the test of any good engineering design: additional modeling resources could be utilized with equal effectiveness in any part of the model.

It would be unrealistic to expect the decision analysis of any large problem to employ all the phases, sensitivity analyses, and models that we have discussed. However, having the concepts and nomenclature necessary to depict these steps is a powerful aid in the planning and execution of a decision analysis. The future should bring continual refinements in the theory and application of the methdology.

Conclusion

The last few years have seen decision analysis grow from a theorist's toy to an important ally of the decision maker. Significant applications have ranged from the desirability of kidney transplants through electric power system planning to the development of policies for space exploration. No one can say when the limits of this revolution will be reached. Whether the limits even exist depends more on man's psychology than on his intellect.

References

[1] R. A. Howard, "Decision analysis: applied decision theory," *Proc. 4th Internat'l Conf. on Operational Research* (Boston, Mass., 1966).
[2] ——, "Bayesian decision models for system engineering," *IEEE Trans. Systems Science and Cybernetics*, vol. SSC-1, pp. 36–40, November 1965.
[3] ——, "Dynamic inference," *Operations Res.*, vol. 13, pp. 712–733, September–October 1965.
[4] ——, "Information value theory," *IEEE Trans. Systems Science and Cybernetics*, vol. SSC-2, pp. 22–26, August 1966.
[5] ——, "Value of information lotteries," *IEEE Trans. Systems Science and Cybernetics*, vol. SSC-3, pp. 54–60, June 1967.

Reprinted from IEEE TRANSACTIONS ON SYSTEMS SCIENCE AND CYBERNETICS
Vol. SSC-4, No. 3, September 1968
pp 211–219
Copyright 1968, and reprinted by permission of the copyright owner
PRINTED IN THE U.S.A.

A Tutorial Introduction to Decision Theory

D. WARNER NORTH

Abstract—Decision theory provides a rational framework for choosing between alternative courses of action when the consequences resulting from this choice are imperfectly known. Two streams of thought serve as the foundations: utility theory and the inductive use of probability theory.

The intent of this paper is to provide a tutorial introduction to this increasingly important area of systems science. The foundations are developed on an axiomatic basis, and a simple example, the "anniversary problem," is used to illustrate decision theory. The concept of the value of information is developed and demonstrated. At times mathematical rigor has been subordinated to provide a clear and readily accessible exposition of the fundamental assumptions and concepts of decision theory. A sampling of the many elegant and rigorous treatments of decision theory is provided among the references.

Introduction

THE NECESSITY of making decisions in the face of uncertainty is an integral part of our lives. We must act without knowing the consequences that will result from the action. This uncomfortable situation is particularly acute for the systems engineer or manager who must make far-reaching decisions on complex issues in a rapidly changing technological environment. Uncertainty appears as the dominant consideration in many systems problems as well as in decisions that we face in our personal lives. To deal with these problems on a rational basis, we must develop a theoretical structure for decision making that includes uncertainty.

Confronting uncertainty is not easy. We naturally try to avoid it, sometimes we even pretend it does not exist. Our primitive ancestors sought to avoid it by consulting soothsayers and oracles who would "reveal" the uncertain future. The methods have changed: astrology and the reading of sheep entrails are somewhat out of fashion today, but predictions of the future still abound. Much current scientific effort goes into forecasting future economic and technological developments. If these predictions are assumed to be completely accurate, the uncertainty in many systems decisions is eliminated. The outcome resulting from a possible course of action may then be presumed to be known. Decision making becomes an optimization problem, and techniques such as mathematical programming may be used to obtain a solution. Such problems may be quite difficult to solve, but this difficulty should not obscure the fact that they represent the limiting case of perfect predictions. It is often tempting to assume perfect predictions, but in so doing we may be eliminating the most important features of the problem.[1] We should like to include in the analysis not just the predictions themselves, but also a measure of the confidence we have in these predictions. A formal theory of decision making must take uncertainty as its departure point and regard precise knowledge of outcomes as a limiting special case.

Before we begin our exposition, we will clarify our point of view. We shall take the engineering rather than the purely scientific viewpoint. We are not observing the way people make decisions; rather we are participants in the decision-making process. Our concern is in actually making a decision, i.e., making a choice between alternative ways of allocating resources. We must assume that at least two distinct alternatives exist (or else there is no element of choice and, consequently, no problem). Alternatives are distinct only if they result in different (uncertain) rewards or penalties for the decision maker; once the decision has been made and the uncertainty resolved, the resource allocation can be changed only by incurring some penalty.

What can we expect of a general theory for decision making under uncertainty? It should provide a framework in which all available information is used to deduce which of the decision alternatives is "best" according to the decision maker's preferences. But choosing an alternative that is consistent with these preferences and present knowledge does not guarantee that we will choose the alternative that by hindsight turns out to be most profitable.

We might distinguish between a good decision and a good *outcome*. We are all familiar with situations in which careful management and extensive planning produced poor results, while a disorganized and badly managed competitor achieved spectacular success. As an extreme example, place yourself in the position of the company president who has discovered that a valuable and trusted subordinate whose past judgment had proved unfailingly accurate actually based his decisions upon the advice of a gypsy fortune teller. Would you promote this man or fire him? The answer, of course, is to fire him and hire the gypsy as a consultant. The availability of such a clairvoyant to provide perfect information would make decision theory unnecessary. But we should not confuse the two. Decision theory is not a substitute for the fortune teller. It is rather a procedure that takes account of all available information to give us the best possible logical

Manuscript received May 8, 1968. An earlier version of this paper was presented at the IEEE Systems Science and Cybernetics Conference, Washington, D.C., October 17, 1966. This research was supported in part by the Graduate Cooperative Fellowship Program of the National Science Foundation at Stanford University, Stanford, Calif.

The author is with the Systems Sciences Area, Stanford Research Institute, Menlo Park, Calif. 94025.

[1] For further discussion of this point, see Howard [10] and Klein and Meckling [14].

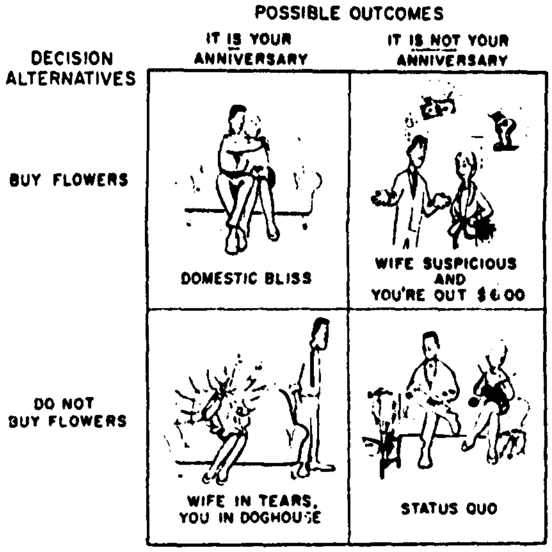


Fig. 1. Anniversary problem payoff matrix.

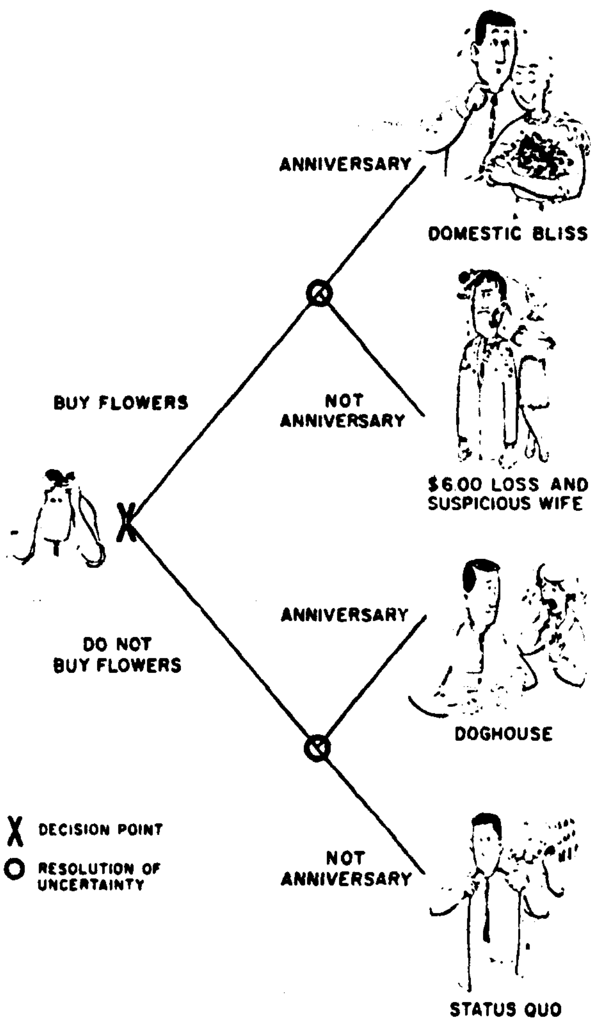


Fig. 2. Diagram of anniversary decision.

decision. It will minimize the consequences of getting an unfavorable outcome, but we cannot expect our theory to shield us from all "bad luck." The best protection we have against a bad outcome is a good decision.

Decision theory may be regarded as a formalization of common sense. Mathematics provides an unambiguous language in which a decision problem may be represented. There are two dimensions to this representation that will presently be described: value, by means of utility theory, and information, by means of probability theory. In this representation, the large and complex problems of systems analysis become conceptually equivalent to simple problems in our daily life that we solve by "common sense." We will use such a problem as an example.

You are driving home from work in the evening when you suddenly recall that your wedding anniversary comes about this time of year. In fact, it seems quite probable (but not certain) that it is today. You can still stop by the florist shop and buy a dozen roses for your wife, or you may go home empty-handed and hope the anniversary date lies somewhere in the future (Fig. 1). If you buy the roses and it is your anniversary, your wife is pleased at what a thoughtful husband you are and your household is the very epitome of domestic bliss. But if it is not your anniversary, you are poorer by the price of the roses and your wife may wonder whether you are trying to make amends for some transgression she does not know about. If you do not buy the roses, you will be in the clear if it is not your anniversary; but if it is, you may expect a temper tantrum from your wife and a two-week sentence to the doghouse. What do you do?

We shall develop the general tools for solving decision problems and then return to this simple example. The reader might consider how he would solve this problem by "common sense" and then compare his reasoning with the formal solution which we shall develop later (Fig. 2).

The Machinery of Decision Making

Utility Theory

The first stage in setting up a structure for decision making is to assign numerical values to the possible outcomes. This task falls within the area covered by the modern theory of utility. There are a number of ways of developing the subject; the path we shall follow is that of Luce and Raiffa [16].[2]

The first and perhaps the biggest assumption to be made is that any two possible outcomes resulting from a decision can be compared. Given any two possible outcomes or prizes, you can say which you prefer. In some cases you might say that they were equally desirable or undesirable, and therefore you are indifferent. For example, you might prefer a week's vacation in Florida to a season ticket to the symphony. The point is not that the vacation costs more than the symphony tickets, but rather

[2] The classical reference on modern utility theory is von Neumann and Morgenstern [22]. A recent survey of the literature on utility theory has been made by Fishburn [5].

that you prefer the vacation. If you were offered the vacation or the symphony tickets on a nonnegotiable basis, you would choose the vacation.

A reasonable extension of the existence of your preference among outcomes is that the preference be transitive; if you prefer A to B and B to C, then it follows that you prefer A to C.[3]

The second assumption, originated by von Neumann and Morgenstern [22], forms the core of modern utility theory: you can assign preferences in the same manner to lotteries involving prizes as you can to the prizes themselves. Let us define what we mean by a lottery. Imagine a pointer that spins in the center of a circle divided into two regions, as shown in Fig. 3. If you spin the pointer and it lands in region I, you get prize A; if it lands in region II, you get prize B. We shall assume that the pointer is spun in such a way that when it stops, it is equally likely to be pointing in any given direction. The fraction of the circumference of the circle in region I will be denoted P, and that in region II as $1 - P$. Then from the assumption that all directions are equally likely, the probability that the lottery gives you prize A is P, and the probability that you get prize B is $1 - P$. We shall denote such a lottery as $(P,A;1 - P,B)$ and represent it by Fig. 4.

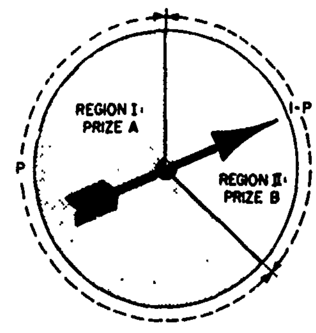


Fig. 3. A lottery.

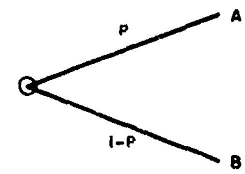


Fig. 4. Lottery diagram.

Now suppose you are asked to state your preferences for prize A, prize B, and a lottery of the above type. Let us assume that you prefer prize A to prize B. Then it would seem natural for you to prefer prize A to the lottery, $(P,A;1 - P,B)$, between prize A and prize B, and to prefer this lottery between prize A and prize B to prize B for all probabilities P between 0 and 1. You would rather have the preferred prize A than the lottery, and you would rather have the lottery than the inferior prize B. Furthermore, it seems natural that, given a choice between two lotteries involving prizes A and B, you would choose the lottery with the higher probability of getting the preferred prize A, i.e., you prefer lottery $(P,A;1 - P,B)$ to $(P',A;1 - P',B)$ if and only if P is greater than P'.

The final assumptions for a theory of utility are not quite so natural and have been the subject of much discussion. Nonetheless, they seem to be the most reasonable basis for logical decision making. The third assumption is that there is no intrinsic reward in lotteries, that is, "no fun in gambling." Let us consider a compound lottery, a lottery in which at least one of the prizes is not an outcome but another lottery among outcomes. For example, consider the lottery $(P,A;1 - P,(P',B;1 - P',C))$. If the pointer of Fig. 3 lands in region I, you get prize A; if it lands in region II, you receive another lottery that has

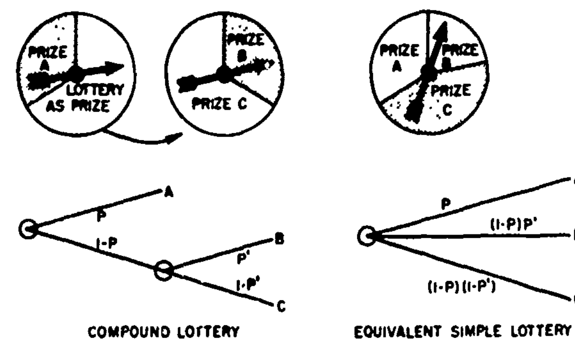


Fig. 5. "No fun in gambling."

different prizes and perhaps a different division of the circle (Fig. 5). If you spin the second pointer you will receive prize B or prize C, depending on where this pointer lands. The assumption is that subdividing region II into two parts whose proportions correspond to the probabilities P' and $1 - P'$ of the second lottery creates an equivalent simple lottery in which all of the prizes are outcomes. According to this third assumption, you can decompose a compound lottery by multiplying the probability of the lottery prize in the first lottery by the probabilities of the individual prizes in the second lottery; you should be indifferent between $(P,A;1 - P,(P',B;1 - P',C))$ and $(P,A;P' - PP',B;1 - P - P' + PP',C)$. In other words, your preferences are not affected by the way in which the uncertainty is resolved—bit by bit, or all at once. There is no value in the lottery itself; it does not matter whether you spin the pointer once or twice.

Fourth, we make a continuity assumption. Consider three prizes, A, B, and C. You prefer A to C, and C to B (and, as we have pointed out, you will therefore prefer A to B). We shall assert that there must exist some probability P so that you are indifferent to receiving prize C or

[3] Suppose not: you would be at least as happy with C as with A. Then if a little man in a shabby overcoat came up and offered you C instead of A, you would presumably accept. Now you have C; and since you prefer B to C, you would presumably pay a sum of money to get B instead. Once you had B, you prefer A; so you would pay the man in the shabby overcoat some more money to get A. But now you are back where you started, with A, and the little man in the shabby overcoat walks away counting your money. Given that you accept a standard of value such as money, transitivity prevents you from becoming a "money pump."

the lottery $(P,A;1 - P,B)$ between A and B. C is called the certain equivalent of the lottery $(P,A;1 - P,B)$, and on the strength of our "no fun in gambling" assumption, we assume that interchanging C and the lottery $(P,A;1 - P,B)$ as prizes in some compound lottery does not change your evaluation of the latter lottery. We have *not* assumed that, given a lottery $(P,A;1 - P,B)$, there exists a Prize C intermediate in value between A and B so that you are indifferent between C and $(P,A;1 - P,B)$. Instead we have assumed the existence of the probability P. Given prize A preferred to prize C preferred to prize B, for some P between 0 and 1, there exists a lottery $(P,A;1 - P,B)$ such that you are indifferent between this lottery and Prize C. Let us regard the circle in Fig. 3 as a "pie" to be cut into two pieces, region I (obtain prize A) and region II (obtain prize B). The assumption is that the "pie" can be divided so that you are indifferent as to whether you receive the lottery or intermediate prize C.

Is this continuity assumption reasonable? Take the following extreme case:

A = receive \$1;
B = death;
C = receive nothing (status quo).

It seems obvious that most of us would agree A is preferred to C, and C is preferred to B; but is there a probability P such that we would risk death for the possibility of gaining \$1? Recall that the probability P can be arbitrarily close to 0 or 1. Obviously, we would not engage in such a lottery with, say, $P = 0.9$, i.e., a 1-in-10 chance of death. But suppose $P = 1 - 1 \times 10^{-50}$, i.e., the probability of death as opposed to \$1 is not 0.1 but 10^{-50}. The latter is considerably less than the probability of being struck on the head by a meteor in the course of going out to pick up a \$1 bill that someone has dropped on your doorstep. Most of us would not hesitate to pick up the bill. Even in this extreme case where death is a prize, we conclude the assumption is reasonable.

We can summarize the assumptions we have made into the following axioms.

A, B, C are prizes or outcomes resulting from a decision.

Notation:

- $>$ means "is preferred to;"
- $A > B$ means A is preferred to B;
- $\sim$ means "is indifferent to;"
- $A \sim B$ means the decision maker is indifferent between A and B.

Utility Axioms:

1) Preferences can be established between prizes and lotteries in an unambiguous fashion. These preferences are transitive, i.e.,

$$A > B,\quad B > C \quad \text{implies } A > C$$
$$A \sim B,\quad B \sim C \quad \text{implies } A \sim C.$$

2) If $A > B$, then $(P,A;1 - P,B) > (P',A;1 - P',B)$ if and only if $P > P'$.

3) $(P,A;1 - P,(P',B;1 - P',C)) \sim (P,A;P' - PP',B;1 - P - P' + PP',C)$, i.e., there is "no fun in gambling."

4) If $A > C > B$, there exists a P with $0 < P < 1$ so that

$$C \sim (P,A;1 - P,B)$$

i.e., it makes no difference to the decision maker whether C or the lottery $(P,A;1 - P,B)$ is offered to him as a prize.

Under these assumptions, there is a concise mathematical representation possible for preferences: a utility function $u(\ )$ that assigns a number to each lottery or prize. This utility function has the following properties:

$$u(A) > u(B) \text{ if and only if } A > B \qquad (1)$$

if $C \sim (P,A;1 - P,B)$,

$$\text{then } u(C) = P \cdot u(A) + (1 - P) \cdot u(B) \qquad (2)$$

i.e., the utility of a lottery is the mathematical expectation of the utility of the prizes. It is this "expected value" property that makes a utility function useful because it allows complicated lotteries to be evaluated quite easily.

It is important to realize that all the utility function does is provide a means of consistently describing the decision maker's preferences through a scale of real numbers, providing these preferences are consistent with the previously mentioned assumptions 1) through 4). The utility function is no more than a means to logical deduction based on given preferences. The preferences come first and the utility function is only a convenient means of describing them. We can apply the utility concept to almost any sort of prizes or outcomes, from battlefield casualties or achievements in space to preferences for Wheaties or Post Toasties. All that is necessary is that the decision maker have unambiguous preferences and be willing to accept the basic assumptions.

In many practical situations however, outcomes are in terms of dollars and cents. What does the utility concept mean here? For an example, let us suppose you were offered the following lottery: a coin will be flipped, and if you guess the outcome correctly, you gain \$100. If you guess incorrectly, you get nothing. We shall assume you feel that the coin has an equal probability of coming up heads or tails; it corresponds to the "lottery" which we have defined in terms of a pointer with $P = 1/2$. How much would you pay for such a lottery? A common answer to this academic question is "up to \$50," the average or expected value of the outcomes. When real money is involved, however, the same people tend to bid considerably lower, the average bid is about \$20.[4] A group of Stanford University graduate students was actually confronted with a \$100 pile of bills and a 1964 silver quarter to flip. The average of the sealed bids for this game was slightly under \$20, and only 4 out of 46 ventured to bid as high as \$40. (The high bidder, at \$45.61, lost and the proceeds were used for a class party.) These results are quite typical; in fact, professional engineers and managers are, if any-

[4] Based on unpublished data obtained by Prof. R. A. Howard of Stanford University, Stanford, Calif.

thing, more conservative in their bids than the less affluent students.

The lesson to be learned here is that, by and large, most people seem to be averse to risk in gambles involving what is to them substantial loss. They are willing to equate the value of a lottery to a sure payoff or certain equivalent substantially less than the expected value of the outcomes. Similarly, most of us are willing to pay more than the expected loss to get out of an unfavorable lottery. This fact forms the basis of the insurance industry.

If you are very wealthy and you are confronted with a small lottery, you might well be indifferent to the risk. An unfavorable outcome would not deplete your resources, and you might reason that you will make up your losses in future lotteries; the "law of averages" will come to your rescue. You then evaluate the lottery at the expected value of the prizes. For example, the (1/2, \$0; 1/2, \$100) lottery would be worth 1/2(\$0) + 1/2(\$100) = \$50 to you. Your utility function is then a straight line, and we say you are an "expected value" decision maker. For lotteries involving small prizes, most individuals and corporations are expected value decision makers. We might regard this as a consequence to the fact that any arbitrary utility curve for money looks like a straight line if we look at a small enough section of it. Only when the prizes are substantial in relation to our resources does the curvature become evident. Then an unfavorable outcome really hurts. For these lotteries most of us become quite risk averse, and expected value decision making does not accurately reflect our true preferences.

Let us now describe one way you might construct your own utility curve for money, say, in the amounts of \$0 to \$100, in addition to your present assets. The utility function is arbitrary as to choice of zero point and of scale factor; changing these factors does not lead to a change in the evaluation of lotteries using properties (1) and (2). Therefore, we can take the utility of \$0 as 0 and the utility of \$100 as 1. Now determine the minimum amount you would accept in place of the lottery of flipping a coin to determine whether you receive \$0 or \$100. Let us say your answer is \$27. Now determine the certain equivalent of the lotteries (1/2, \$0; 1/2, \$27), and (1/2, \$27; 1/2, \$100), and so forth. We might arrive at a curve like that shown in Fig. 6.

We have simply used the expected value property (2) to construct a utility curve. This same curve, however, allows us to use the same expected utility theorem to evaluate new lotteries; for example, (1/2, \$30; 1/2 \$80). From Fig. 6, $u(\$30) = 0.54$, $u(\$80) = 0.91$, and therefore $1/2\ u(\$30) + 1/2\ u(\$80) = u(x) \rightarrow x = \$49$. If you are going to be consistent with the preferences you expressed in developing the utility curve, you will be indifferent between \$49 and this lottery. Moreover, this amount could have been determined from your utility curve by a subordinate or perhaps a computer program. You could send your agent to make decisions on lotteries by using your utility curve, and he would make them to reflect *your* preference for amounts in the range \$0 to \$100.

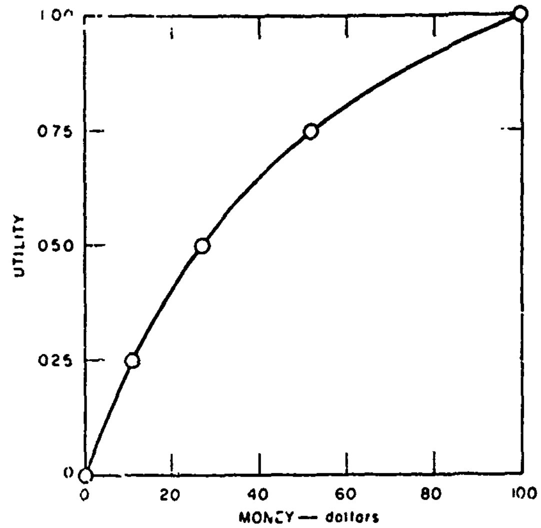


Fig. 6. Utility curve for money: \$0 to \$100.

Even without such a monetary representation, we can always construct a utility function on a finite set of outcomes by using the expected value property (2). Let us choose two outcomes, one of which is preferred to the other. If we set the utilities arbitrarily at 1 for the preferred outcome and 0 for the other, we can use the expected value property (2) of the utility function to determine the utility of the other prizes. This procedure will always work so long as our preferences obey the axioms, but it may be unwieldy in practice because we are asking the decision maker to assess simultaneously his values in the absence of uncertainty and his preference among risks. The value of some outcome is accessible only by reference to a lottery involving the two "reference" outcomes. For example, the reference outcomes in the anniversary problem might be "domestic bliss" = 1 and "doghouse" = 0. We could then determine the utility of "status quo" as 0.91 since the husband is indifferent between the outcome "status quo" and a lottery in which the chances are 10 to 1 of "domestic bliss" as opposed to the "doghouse." Similarly, we might discover that a utility of 0.667 should be assigned to "suspicious wife and \$6 wasted on roses," since our friend is indifferent between this eventuality and a lottery in which the probabilities are 0.333 of "doghouse" and 0.667 of "domestic bliss." Of course, to be consistent with the axioms, our friend must be indifferent between "suspicious wife, etc.," and a 0.73 probability of "status quo" and a 0.27 probability of "doghouse." If the example included additional outcomes as well, he might find it quite difficult to express his preferences among the lotteries in a manner consistent with the axioms. It may be advisable to proceed in two stages; first, a numerical determination of value in a risk-free situation, and then an adjustment to this scale to include preference toward risk.

Equivalent to our first assumption, the existence of transitive preferences, is the existence of some scale of value by which outcomes may be ranked; A is preferred to B if and only if A is higher in value than B. The numerical

structure we give to this value is not important since a monotonic transformation to a new scale preserves the ranking of outcomes that corresponds to the original preferences. No matter what scale of value we use, we can construct a utility function on it by using the expected value property (2), so long as our four assumptions hold. We may as well use a standard of value that is reasonably intuitive, and in most situations money is a convenient standard of economic value. We can then find a monetary equivalent for each outcome by determining the point at which the decision maker is indifferent between receiving the outcome and receiving (or paying out) this amount of money. In addition to conceptual simplicity, this procedure makes it easy to evaluate new outcomes by providing an intuitive scale of values. Such a scale will become necessary later on if we are to consider the value of resolving uncertainty.

We will return to the anniversary decision and demonstrate how this two-step value determination procedure may be applied. But first let us describe how we shall quantify uncertainty.

The Inductive Use of Probability Theory

We now wish to leave the problem of the evaluation of outcomes resulting from a decision and turn our attention to a means of encoding the information we have as to which outcome is likely to occur. Let us look at the limiting case where a decision results in a certain outcome. We might represent an outcome, or an event, which is certain to occur by 1, and an event which cannot occur by 0. A certain event, together with another certain event, is certain to occur; but a certain event, together with an impossible event, is certain not to occur. Most engineers would recognize the aforementioned as simple Boolean equations: $1 \cdot 1 = 1$, $1 \cdot 0 = 0$. Boolean algebra allows us to make complex calculations with statements that may take on only the logical values "true" and "false." The whole field of digital computers is, of course, based on this branch of mathematics.

But how do we handle the logical "maybe?" Take the statement, "It will rain this afternoon." We cannot now assign this statement a logical value of true or false, but we certainly have some feelings on the matter, and we may even have to make a decision based on the truth of the statement, such as whether to go to the beach. Ideally, we would like to generalize the inductive logic of Boolean algebra to include uncertainty. We would like to be able to assign to a statement or an event a value that is a measure of its uncertainty. This value would lie in the range from 0 to 1. A value of 1 indicates that the statement is true or that the event is certain to occur; a value of 0 indicates that the statement is false or that the event cannot occur. We might add two obvious assumptions. We want the value assignments to be unambiguous, and we want the value assignments to be independent of any assumptions that have not been explicitly introduced. In particular, the value of the statement should depend on its content, not on the way it is presented. For example, "It will rain this morning or it will rain this afternoon," should have the same value as "It will rain today."

These assumptions are equivalent to the assertion that there is a function P that gives values between 0 and 1 to events ("the statement is true" is an event) and that obeys the following probability axioms.[5]

Let E and F be events or outcomes that could result from a decision:

1) $P(E) \geq 0$ for any event E;
2) $P(E) = 1$, if E is certain to occur;
3) $P(E \text{ or } F) = P(E) + P(F)$ if E and F are mutually exclusive events (i.e., only one of them can occur).

E *or* F means the event that either E or F occurs. We are in luck. Our axioms are identical to the axioms that form the modern basis of the theory of probability. Thus we may use the whole machinery of probability theory for inductive reasoning.

Where do we obtain the values $P(E)$ that we will assign to the uncertainty of the event E? We get them from our own minds. They reflect our best judgment on the basis of all the information that is presently available to us. The use of probability theory as a tool of inductive reasoning goes back to the beginnings of probability theory. In Napoleon's time, Laplace wrote the following as a part of his introduction to *A Philosophical Essay on Probabilities* ([15], p. 1):

> Strictly speaking it may even be said that nearly all our knowledge is problematical; and in the small numbers of things which we are able to know with certainty, even in the mathematical sciences themselves, the principal means for ascertaining truth—induction and analogy—are themselves based on probabilities....

Unfortunately, in the years following Laplace, his writings were misinterpreted and fell into disfavor. A definition of probability based on frequency came into vogue, and the pendulum is only now beginning to swing back. A great many modern probabilists look on the probability assigned to an event as the limiting fraction of the number of times an event occurred in a large number of independent repeated trials. We shall not enter into a discussion of the general merits of this viewpoint on probability theory. Suffice it to say that the situation is a rare one in which you can observe a great many independent identical trials in order to assign a probability. In fact, in decision theory we are often interested in events that will occur just once. For us, a probability assessment is made on the basis of a state of mind; it is not a property of physical objects to be measured like length, weight, or temperature. When we assign the probability of 0.5 to a coin coming up heads, or equal probabilities to all possible orientations of a pointer, we may be reasoning on the basis of the symmetry of the

[5] Axioms 1) and 2) are obvious, and 3) results from the assumption of invariance to the form of data presentation (the last sentence in the preceding paragraph). Formal developments may be found in Cox [3], Jaynes [12], or Jeffreys [13]. A joint axiomatization of both probability and utility theory has been developed by Savage [20].

physical object. There is no reason to suppose that one side of the coin will be favored over the other. But the physical symmetry of the coin does not lead immediately to a probability assignment of 0.5 for heads. For example, consider a coin that is placed on a drum head. The drum head is struck, and the coin bounces into the air. Will it land heads up half of the time? We might expect that the probability of heads would depend on which side of the coin was up initially, how hard the drum was hit, and so forth. The probability of heads is not a physical parameter of the coin; we have to specify the flipping system as well. But if we knew exactly how the coin were to be flipped, we could calculate from the laws of mechanics whether it would land heads or tails. Probability enters as a means of describing our feelings about the likelihood of heads when our knowledge of the flipping system is not exact. We must conclude that the probability assignment depends on our present state of knowledge.

The most important consequence of this assertion is that probabilities are subject to change as our information improves. In fact, it even makes sense to talk about probabilities of probabilities. A few years ago we might have assigned the value 0.5 to the probability that the surface of the moon is covered by a thick layer of dust. At the time, we might have said, "We are 90 percent certain that our probability assignment after the first successful Surveyor probe will be less than 0.01 or greater than 0.99. We expect that our uncertainty about the composition of the moon's surface will be largely resolved."

Let us conclude our discussion of probability theory with an example that will introduce the means by which probability distributions are modified to include new information: Bayes' rule. We shall also introduce a useful notation. We have stressed that all of our probability assignments are going to reflect a state of information in the mind of the decision maker, and our notation shall indicate this state of information explicitly.

Let A be an event, and let x be a quantity about which we are uncertain; e.g., x is a random variable. The values that x may assume may be discrete (i.e., heads or tails) or continuous (i.e., the time an electronic component will run before it fails). We shall denote by $\{A|S\}$ the probability assigned to the event A on the basis of a state of information S, and by $\{x|S\}$ the probability that the random variable assumes the value x, i.e., the probability mass function for a discrete random variable or the probability density function for a continuous random variable, given a state of information S. If there is confusion between the random variable and its value, we shall write $\{x = x_0|S\}$, where x denotes the random variable and x_0 the value. We shall assume the random variable takes on some value, so the probabilities must sum to 1:

$$\int_x \{x|S\} = 1. \qquad (3)$$

$\int$ is a generalized summation operator representing summation over all discrete values or integration over all continuous values of the random variable. The expected value, or the average of the random variable over its probability distribution, is

$$\langle x|S\rangle = \int_x x\{x|S\}. \qquad (4)$$

One special state of information will be used over and over again, so we shall need a special name for it. This is the information that we now possess on the basis of our prior knowledge and experience, before we have done any special experimenting or sampling to reduce our uncertainty. The probability distribution that we assign to values of an uncertain quantity on the basis of this prior state of information (denoted $\mathcal{E}$) will be referred to as the "prior distribution" or simply the "prior."

Now let us consider a problem. Most of us take as axiomatic the assignment of 0.5 to the probability of heads on the flip of a coin. Suppose we flip thumbtacks. If the thumbtack lands with the head up and point down, we shall denote the outcome of the flip as "heads." If it lands with the head down and the point up, we shall denote the outcome as "tails." The question which we must answer is, "What is p, the probability of heads in flipping a thumbtack?" We will assume that both thumbtack and means of flipping are sufficiently standardized so that we may expect that all flips are independent and have the same probability for coming up heads. (Formally, the flips are Bernoulli trials.) Then the long-run fraction of heads may be expected to approach p, a well-defined number that at the moment we do not know.

Let us assign a probability distribution to this uncertain parameter p. We are all familiar with thumbtacks; we have no doubt dropped a few on the floor. Perhaps we have some experience with spilled carpet tacks, or coin flipping, or the physics of falling bodies that we believe is relevant. We want to encode all of this prior information into the form of a probability distribution on p.

This task is accomplished by using the cumulative distribution function, $\{p \le p_0|\mathcal{E}\}$, the probability that the parameter p will be less than or equal to some specific value of the parameter p_0. It may be convenient to use the complementary cumulative

$$\{p > p_0|\mathcal{E}\} = 1 - \{p \le p_0|\mathcal{E}\}$$

and ask questions such as, "What is the probability that p is greater than $p_0 = 0.5$?"

To make the situation easier to visualize, let us introduce Sam, the neighborhood bookie. We shall suppose that we are forced to do business with Sam. For some value p_0 between 0 and 1, Sam offers us two packages:

Package 1: If measurement of the long run fraction of heads p shows that the quantity is less than or equal to p_0, then Sam pays us $1. If $p > p_0$, then we pay Sam $1.

Package 2: We divide a circle into two regions (as shown in Fig. 3). Region I is defined by a fraction P of the circumference of the circle, and the remainder of the circle constitutes region II. Now a pointer is spun in such a way that when it stops, it is equally likely to be pointing in any

given direction. If the pointer stops in region I, Sam pays us \$1; if it lands in region II, we pay Sam \$1.

Sam lets us choose the fraction P in Package 2, but then he chooses which package we are to receive. Depending on the value of p_0, these packages may be more or less attractive to us, but it is the relative rather than the absolute value of the two packages that is of interest. If we set P to be large, we might expect that Sam will choose package 1, whereas if P is small enough, Sam will certainly choose package 2. Sam wishes (just as we do) to have the package with the higher probability of winning \$1. (Recall this is our second utility axiom.) We shall assume Sam has the same information about thumbtacks that we do, so his probability assignments will be the same as ours. The assumption (utility axiom 4)) is that given p_0, we can find a P such that Packages 1 and 2 represent equivalent lotteries, so $P = \{p \leq p_0|\mathcal{E}\}$.[6] The approach is similar to the well-known method of dividing an extra dessert between two small boys: let one divide and the other choose. The first is motivated to make the division as even as possible so that he will be indifferent as to which half he receives.

Suppose Sam starts at a value $p_0 = 0.5$. We might reason that since nails always fall on the side (heads), and a thumbtack is intermediate between a coin and a nail heads is the more likely orientation; but we are not too sure; we have seen a lot of thumbtacks come up tails. After some thought, we decide that we are indifferent about which package we get if the fraction P is 0.3, so $\{p \leq 0.5|\mathcal{E}\} = 0.30$.

Sam takes other values besides 0.5, skipping around in a random fashion, i.e., 0.3, 0.9, 0.1, 0.45, 0.8, 0.6, etc. The curve that results from the interrogation might look like that shown in Fig. 7. By his method of randomly skipping around, Sam has eliminated any bias in our true feelings that resulted from an unconscious desire to give answers consistent with previous points. In this fashion, Sam has helped us to establish our prior distribution on the parameter p. We may derive a probability density function by taking the derivative of the cumulative distribution function (Fig. 8): $\{p|\mathcal{E}\} = (d/dp_0)\{p \leq p_0|\mathcal{E}\}$.

Now supposing we are allowed to flip the thumbtack 20 times and we obtain 5 heads and 15 tails. How do we take account of this new data in assigning a probability distribution based on the new state of information, which we denote as $\mathcal{E}, E$: our prior experience $\mathcal{E}$ plus E, the 20-flip experiment? We will use one of the oldest (1763) results of probability theory, Bayes' rule. Consider the prior probability that p will take on a specific value and the 20-flip experiment E will have a certain specific outcome (for example, $p = 0.43$; E = 5 heads, 15 tails). Now we can write this joint probability in two ways:

$$\{p,E|\mathcal{E}\} = \{p|E,\mathcal{E}\}\{E|\mathcal{E}\} \tag{5}$$

[6] We have equated the subjective probability that summarised our information about thumbtacks to the more intuitive notion of probability based on symmetry (in Package 2). Such a two-step approach to probability theory has been discussed theoretically by Anscombe and Aumann [1].

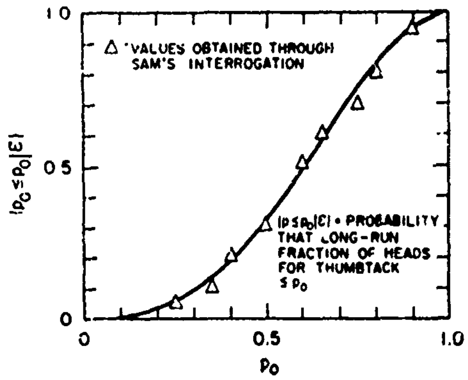


Fig. 7. Cumulative distribution function for thumbtack flipping.

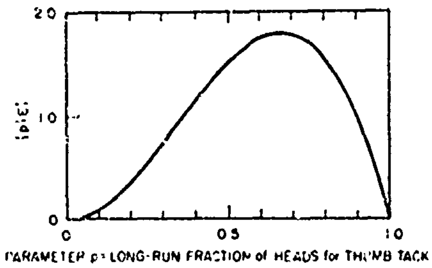


Fig. 8. Prior probability density function.

i.e., as the product of the probability we assign to the experimental outcome E times the probability we would assign to the value of p after we knew the experimental outcome E in addition to our prior information; or

$$\{p,E|\mathcal{E}\} = \{E|p,\mathcal{E}\}\{p|\mathcal{E}\} \tag{6}$$

i.e., the product of the probability of that experimental outcome if we knew that p were the probability of getting heads times our prior probability assessment that p actually takes on that value.

We assumed that probabilities were unambiguous, so we equate these two expressions. Providing $\{E|\mathcal{E}\} \neq 0$, i.e., the experimental outcome is not impossible, we obtain the posterior (after the experiment) probability distribution on p

$$\{p|E,\mathcal{E}\} = \frac{\{E|p,\mathcal{E}\}\{p|\mathcal{E}\}}{\{E|\mathcal{E}\}}. \tag{7}$$

This expression is the well-known Bayes' rule.

$\{E|\mathcal{E}\}$ is the "pre-posterior" probability of the outcome E. It does not depend on p, so it becomes a normalizing factor for the posterior probability distribution. $\{E|p,\mathcal{E}\}$ is the probability of the outcome E if we knew the value p for the probability of heads. This probability is a function of p, usually referred to as the "likelihood function." We notice since p must take on some value, the expectation of the likelihood function over the values of p gives the pre-posterior probability of the experimental outcome:

$$\{E|\mathcal{E}\} = \int_p \{E|p,\mathcal{E}\}\{p|\mathcal{E}\}. \tag{8}$$

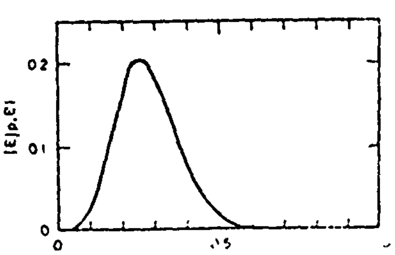


Fig. 9. Likelihood function for 5 heads in 20 trials.

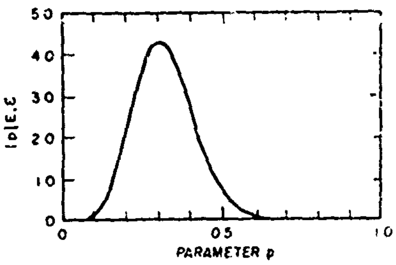


Fig. 10. Posterior probability density function.

Reproduced from best available copy.

For the specific case we are treating, the likelihood function is the familiar result from elementary probability theory for r successes in n Bernoulli trials when the probability of a success is p:

$$\{E|p,\mathcal{E}\} = \frac{n!}{r!(n-r)!}\, p^r (1-p)^{n-r}. \qquad (9)$$

This function is graphed for $r = 5$ heads in $n = 20$ trials in Fig. 9. Multiplying it by the prior $\{p|\mathcal{E}\}$ (Fig. 8) and normalizing by dividing by $\{E|\mathcal{E}\}$ gives us the posterior distribution $\{p|E,\mathcal{E}\}$ (Fig. 10). In this way, Bayes' rule gives us a general means of revising our probability assessments to take account of new information.[7]

Solution of Decision Problems

Now that we have the proper tools, utility theory and probability theory, we return to the anniversary decision problem. We ask the husband, our decision maker, to assign monetary values to the four possible outcomes. He does so as follows:

Domestic bliss	(flowers + anniversary):	\$100
Doghouse	(no flowers, anniversary):	\$ 0
Status quo	(no flowers, no anniversary):	\$ 80
Suspicious wife	(flowers, no anniversary):	\$ 42.

(For example, he is indifferent between "status quo" and "doghouse" provided in the latter case he receives \$80.) His preference for risk is reflected by the utility function of Fig. 6, and he decides that a probability assessment of 0.2 sums up his uncertainty about the possibility of today being his anniversary: the odds are 4 to 1 that it is not his anniversary. Now let us look at the two lotteries that represent his decision alternatives. If he buys the flowers, he has a 0.2 probability of "domestic bliss" and an 0.8 probability of "suspicious wife." The expected utility of the lottery is $0.2(1.0) + 0.8(0.667) = 0.734 = u(\$50)$. On the other hand, if he does not buy the flowers, he has an 0.8 chance of "status quo" and a 0.2 chance of "doghouse." The expected utility of this alternative is $0.8(0.91) + 0.2(0) = 0.728 = u(\$49)$. The first alternative has a slightly higher value to him so he should buy the flowers. On the basis of his values, his risk preference, and his judgment about the uncertainty, buying the flowers is his best alternative. If he were an expected value decision maker, the first lottery would be worth $0.2(\$100) + 0.8(\$42) = \$53.60$ and the second $0.2(0) + 0.8(\$80) = \$64$. In this case he should not buy the flowers.

The foregoing example is, of course, very trivial, but conceptually any decision problem is exactly the same. There is only one additional feature that we may typically expect: in general, decision problems may involve a sequence of decisions. First, a decision is made and then an uncertain outcome is observed; after which another decision is made, and an outcome observed, etc. For example, the decision to develop a new product might go as follows. A decision is made as to whether or not a product should be developed. If the decision is affirmative, an uncertain research and development cost will be incurred. At this point, a decision is made as to whether to go into production. The production cost is uncertain. After the production cost is known, a sale price is set. Finally, the uncertain sales volume determines the profit or loss on the product.

We can handle this problem in the same way as the anniversary problem: assign values to the final outcomes, and probabilities to the various uncertain outcomes that will result from the adoption of a decision alternative. We can represent the problem as a decision tree (Fig. 11), and the solution is conceptually easy. Start at the final outcome, sales volume (the ends of the tree). Go in to the first decision, the sales price (the last to be made chronologically). Compute the utility of the decision alternatives, and choose the one with the highest value. This value becomes the utility of the chance outcome leading to that decision (e.g., production cost). The corresponding certain equivalent in dollars reflects the expected utility of reaching that point in the tree. In this fashion, we work backwards to the start of the tree, finding the best decision alternatives and their values at each step.

Many decision problems encountered in actual practice are extremely complex, and a decision tree approach may not always be appropriate. If all quantities concerned in the problem were considered uncertain (with prior distributions), the problem might be computationally intractable. It is often advisable to solve the model deterministically as a first approximation. We approximate all uncertain quantities with a single best estimate and then examine the decision; i.e., if research and development costs, production costs, and sales volume took the

[7] For certain sampling processes having special statistical properties, assumption of a prior probability distribution from a particular family of functions leads to a simple form for Bayes' rule. An extensive development of this idea of "conjugate distributions" has been accomplished by Raiffa and Schlaifer [19].

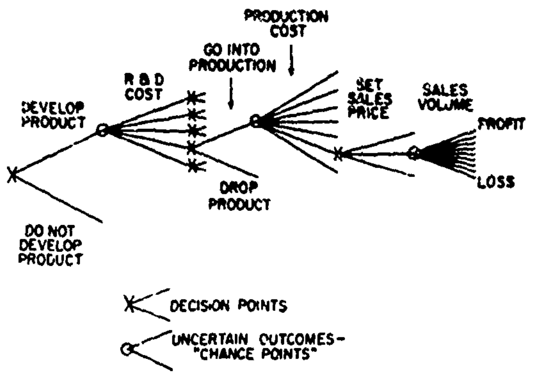


Fig. 11. Product development decision tree.

values we consider most likely, would it then be advisable to develop the product? This deterministic phase will usually give us some insight into the decision. Moreover, we can perform a sensitivity analysis by varying quantities that we believe are uncertain to determine how they affect the decision. The decision may be quite insensitive to some quantities, and these quantities may be treated as certain (uncertainty is neglected if it appears not to affect the decision). On the other hand, if a variation that lies within the range of uncertainty of a factor causes a major shift in the decision (i.e., from "develop the product" to "do not develop the product"), we shall certainly wish to encode our feelings about the uncertainty of that quantity by a prior distribution.[6]

The Value of Resolving Uncertainties

There is a class of alternatives usually available to the decision maker that we have not yet mentioned: activities that allow him to gather more information to diminish the uncertainties before he makes the decision. We have already seen how new information may be incorporated into probability assessments through Bayes' rule and we noted that we can assign a probability distribution to the results of the information gathering by means of the pre-posterior probability distribution. Typical information-gathering activities might include market surveys, pilot studies, prototype construction, test marketing, or consulting with experts. These activities invariably cost the decision maker time and resources; he must pay a price for resolving uncertainty.

Let us return to the husband with the anniversary problem. Suppose he has the option of calling his secretary. If it is his anniversary, his secretary will certainly tell him. But if it is not, she may decide to play a trick and tell him that today is his anniversary. He assigns probability 0.5 to such practical joking. In any event, the secretary will spread the word around the office and our friend will get some good natured heckling, which he views as having a value of minus $10.

[6] The decision analysis procedure has been described in detail by Howard [8].

How will the secretary's information change his assessment of the probability that today is his anniversary? If she says, "No, it is not your anniversary," he may be sure that it is not; but if she says "Yes, it is," she could be joking. We can compute the new assessment of the probability from Bayes' rule. This new probability is equal to the probability 0.2 that she says yes and it really is his anniversary, divided by his prior estimate, $0.2 + 0.5 \times 0.8 = 0.6$, that she will say yes regardless of the date of his anniversary. Hence the probability assignment revised to include the secretary's yes answer is 0.333.

What is the value of this new alternative to our friend? If his secretary says no (probability 0.4), he may return home empty-handed and be assured of "status quo." On the other hand, if she says yes (probability 0.6), he will buy the flowers. In either case, he has incurred a cost of $10 which must be subtracted from the values of the outcomes. Calling the secretary then has a utility of

$$0.4\,u(\$70) + 0.6\,[0.333\,u(\$90) + 0.667\,u(\$32)] = 0.344 + 0.416 = 0.760 = u(\$53.50).$$

Since this value of $53.50 exceeds the value of $50 for his previous best alternative (buy flowers), our friend should call his secretary. If the husband were an expected value decision maker, the alternative of calling the secretary would have a value of

$$0.4\,(\$70) + 0.6\,[0.333\,(\$90) + 0.667\,(\$32)] = \$58.80$$

which is less than the value of $64 for the "do not buy flowers" alternative; in this case our friend should not call his secretary. It is evident that in this example preference toward risk is very important in determining the decision maker's best course of action.

In the complex decision problems normally encountered in practice, there are usually several alternative options available for diminishing the uncertainty associated with the unknown factors. In theory, the expected gain for each type of sampling could be computed and compared with the cost of sampling as we have just done in the simple anniversary example. But these calculations can be quite involved as a rule, and there may be a great many alternative ways of gathering information. Often the relevant questions are, first, "Should we sample at all?" and then, "What kind of sampling is best for us?"

It is often useful to look at the limiting case of complete resolution of uncertainty, which we call perfect information. We can imagine that a gypsy fortune teller who always makes correct predictions is, in fact, available to us. The value of perfect information is the amount that we are willing to pay her to tell us exactly what the uncertain quantity will turn out to be. Note that her answer may be of little value to us—we may be planning to take the best decision alternative already. On the other hand, her perfect information may be quite valuable; it may allow us to avoid an unfavorable outcome. We are going to have to pay her before we hear her information; our payment will reflect what we expect the information to be on the basis of our prior probability assessment.

In the husband's anniversary problem, perfect information might correspond to a secretary who is certain to tell him if today is his anniversary. If he could act on this information, he would buy flowers if it were his anniversary and would not buy flowers otherwise. Since he feels that there is a 0.2 chance the secretary will tell him that it is his anniversary, the expected utility of the outcomes if he bases his decision on perfect information is $0.2\,u(\$100 - b) + 0.8\,u(\$80 - b)$ where b is the amount he must pay to get the information. By setting this expression equal to 0.734, the expected utility of his best alternative based on prior information, we can solve for $b = \$33.50$. The husband should consider for more detailed analysis only those opportunities for resolving his uncertainty that "cost" him $33.50 or less. If he were an expected value decision maker, perfect information would be of less value to him; he would be willing to pay a maximum of only $20 for it.[9]

Summary

Decision theory is a way of formalizing common sense. The decision maker analyzes the possible outcomes resulting from his available alternatives in two dimensions: value (by means of utility theory) and probability of occurrence. He then chooses the alternative that he expects to have the highest value. He cannot guarantee that the outcome will be as good as he might hope for, but he has made the best decision he can, based on his preferences and available knowledge. Inference using Bayes' rule allows the decision maker to evaluate information gathering activities that will reduce his uncertainty.

Decision theory gives no magical formulas for correct decisions. In fact, it forces the decision maker to rely more strongly than ever on his own preferences and judgments. But it does give him a logical framework in which to work, a framework that is adaptable in principle to all decision problems, from the simplest to the most complex. As modern society continues to grow in size and complexity, such a framework for decision making will become more and more necessary.

[9] Additional discussion regarding the value of information in decision theory is available from many sources, most notably Howard [8b], [9], [11] and Raiffa and Schlaifer [19].

Acknowledgment

The author is deeply indebted to Prof. R. A. Howard for valuable guidance and advice. Much of the viewpoint presented in this paper was acquired as a result of two years of assisting Prof. Howard in his course in Decision Analysis at Stanford University, Stanford, Calif.

References

[1] F. J. Anscombe and R. J. Aumann, "A definition of subjective probability," *Ann. Math. Statist.*, vol. 34, pp. 199–205, 1963.
[2] T. Bayes, "An essay toward solving a problem in the doctrine of chances," *Phil. Trans. Roy. Soc.* (London), vol. 53, pp. 370–418, 1763.
[3] R. T. Cox, *The Algebra of Probable Inference.* Baltimore, Md.: The John Hopkins Press, 1961.
[4] P. C. Fishburn, *Decision and Value Theory.* New York: Wiley, 1964.
[5] ——, "Utility theory," *Management Sci.*, vol. 14, pp. 335–378, January 1968.
[6] ——, "On the prospects of a useful unified theory of value for engineering," *IEEE Trans. Systems Science and Cybernetics*, vol. SSC-2, pp. 27–35, August 1966.
[7] R. A. Howard, "Bayesian decision models for system engineering," *IEEE Trans. Systems Science and Cybernetics*, vol. SSC-1, pp. 36–40, November 1965.
[8] a) ——, "Decision analysis: applied decision theory," *Proc. 4th Internat'l Conf. Operational Res.* (Boston, Mass., 1966).
b) ——," The foundations of decision analysis" this issue, pp. 211–219.
[9] ——, "Information value theory," *IEEE Trans. Systems Science and Cybernetics*, vol. SSC-2, pp. 22–26, August 1966.
[10] ——, "The science of decision-making" (unpublished).
[11] ——, "Value of information lotteries," *IEEE Trans. Systems Science and Cybernetics*, vol. SSC-3, pp. 54–60, June 1967.
[12] E. T. Jaynes, "Probability theory in science and engineering," Field Research Laboratory, Socony Mobil Oil Co., Inc., Dallas, Tex., 1959.
[13] H. Jeffreys, *Theory of Probability*, 3rd ed. Oxford: Clarendon Press, 1961.
[14] B. Klein and W. Meckling, "Applications of operations research to development decisions," *Operations Res.*, vol. 6, pp. 352–363, May–June 1958.
[15] P. S. Laplace, *Essai Philosophique sur les Probabilités* (Paris, 1814), translation. New York: Dover, 1951.
[16] R. D. Luce and H. Raiffa, *Games and Decisions: Introduction and Critical Survey.* New York: Wiley, 1957.
[17] J. W. Pratt, H. Raiffa, and R. Schlaifer, "The foundations of decision under uncertainty: an elementary exposition," *J. Am. Statist. Assoc.*, vol. 59, pp. 353–375, June 1964.
[18] ——, *Introduction to Statistical Decision Theory.* New York: McGraw-Hill, 1965.
[19] H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory.* Boston, Mass.: Graduate School of Business, Harvard University, 1961.
[20] L. J. Savage, *The Foundations of Statistics.* New York: Wiley, 1954.
[21] R. Schlaifer, *Probability and Statistics for Business Decisions.* New York: McGraw-Hill, 1959.
[22] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, 2nd ed. Princeton, N.J.: Princeton University Press, 1947.

Reprinted from IEEE TRANSACTIONS ON SYSTEMS SCIENCE AND CYBERNETICS
Vol. SSC-4, No. 3, September 1968
pp. 200–210
Copyright 1968, and reprinted by permission of the copyright owner
PRINTED IN THE U.S.A.

Reprinted from Proceedings of the Fourth International Conference on Operational Research (D.B. Hertz and J. Melese, Eds.). New York: Wiley, 1968. Pp. 55-71.

DECISION ANALYSIS: APPLIED DECISION THEORY

Analyse des Décisions: Théorie Appliquée des Décisions

RONALD A. HOWARD

Institute in Engineering-Economic Systems
Stanford University, California
United States of America

1. INTRODUCTION

Decision theory in the modern sense has existed for more than a decade. Most of the effort among the present developers of the theory has been devoted to Bayesian analysis of problems formerly treated by classical statistics. Many practical management decision problems, however, can be handled by formal structures that are far from novel theoretically. The world of top management decision making is not often structured by simple Bernoulli, Poisson, or normal models.

Indeed, Bayes's theorem itself may not be so important. A statistician for a major company wrote a report in which he commented that for all the talk about the Bayesian revolution he did not know of a single application in the company in which Bayes's theorem was actually used. The observation was probably quite correct—but what it shows by implication is that the most significant part of the revolution is not Bayes's theorem or conjugate distributions but rather the concept of probability as a state of mind, a 200-year-old concept. Thus the real promise of decision theory lies in its ability to provide a broad logical basis for decision making in the face of uncertainty rather than in any specific models.

The purpose of this article is to outline a formal procedure for the analysis of decision problems, a procedure that I call "decision analysis." We shall also discuss several of the practical problems that arise when we attempt to apply the decision analysis formalism.

2. DECISION ANALYSIS

To describe decision analysis it is first necessary to define a decision. A decision is an irrevocable allocation of resources, irrevocable in the sense that it is impossible or extremely costly to change back to the situation that existed before making the decision. Thus for our purposes a decision is not a mental commitment to follow a course of action but rather the actual pursuit of that course of action. This definition often serves to identify the real decision maker within a loosely structured organization. Finding the exact nature of the decision to be

Preceding page blank

made, however, and who will make it, remains one of the fundamental problems of the decision analyst.

Having defined a decision, let us clarify the concept by drawing a necessary distinction between a good decision and a good outcome. A good decision is a logical decision—one based on the uncertainties, values, and preferences of the decision maker. A good outcome is one that is profitable or otherwise highly valued. In short, a good outcome is one that we wish would happen. Hopefully, by making good decisions in all the situations that face us we shall ensure as high a percentage as possible of good outcomes. We may be disappointed to find that a good decision has produced a bad outcome or dismayed to learn that someone who has made what we consider to be a bad decision has enjoyed a good outcome. Yet, pending the invention of the true clairvoyant, we find no better alternative in the pursuit of good outcomes than to make good decisions.

Decision analysis is a logical procedure for the balancing of the factors that influence a decision. The procedure incorporates uncertainties, values, and preferences in a basic structure that models the decision. Typically, it includes technical, marketing, competitive, and environmental factors. The essence of the procedure is the construction of a structural model of the decision in a form suitable for computation and manipulation; the realization of this model is often a set of computer programs.

2.1. The Decision Analysis Procedure

Table 1 lists the three phases of a decision analysis that are worth distinction: the deterministic, probabilistic, and post-mortem phases.

TABLE 1

The Decision Analysis Procedure

I. Deterministic phase
   1. Define the decision
   2. Identify the alternatives
   3. Assign values to outcomes
   4. Select state variables
   5. Establish relationship at state variables
   6. Specify time preference

   Analysis: (a) Determine dominance to eliminate alternatives
   (b) Measure sensitivity to identify crucial state variables

II. Probabilistic phase
   1. Encode uncertainty on crucial state variables
      Analysis: Develop profit lottery
   2. Encode risk preference
      Analysis: Select best alternative

III. Post-mortem phase

   Analysis: (a) Determine value of eliminating uncertainty in crucial state variables
   (b) Develop most economical information-gathering program

2.1.1. *The Deterministic Phase*

The first step in the deterministic phase is to answer the question, "What decision must be made?" Strange as it may seem, many people with what appear to be decision problems have never asked themselves that question. We must distinguish between situations in which there is a decision to be made and situations in which we are simply worried about a bad outcome. If we have resources to allocate, we have a decision problem, but if we are only hand wringing about circumstances beyond our control no formal analysis will help. The difference is that between selecting a surgeon to operate on a member of your family and waiting for the result of the operation. We may be in a state of anguish throughout, but decision analysis can help only with the first question.

The next step is to identify the alternatives that are available, to answer the question, "What courses of action are open to us?" Alternative generation is the most creative part of the decision analysis procedure. Often the introduction of a new alternative eliminates the need for further formal analysis. Although the synthesis of new alternatives necessarily does not fall within the province of the decision analysis procedure, the procedure does evaluate alternatives and thereby suggests the defects in present alternatives that new alternatives might remedy. Thus the existence of an analytic procedure is the first step toward synthesis.

We continue the deterministic phase by assigning values to the various outcomes that might be produced by each alternative. We thus answer the question, "How are you going to determine which outcomes are good and which are bad?" In business problems this will typically be a measure of profit. Military and governmental applications should also consider profit, measured perhaps with more difficulty, because these decision makers are also allocating the economic resources of the nation. Even when we agree on the measure of profit to be assigned to each outcome, it may be difficult to make the assignment until the values of a number of variables associated with each outcome are specified. We call these variables the state variables of the decision. Their selection is the next step in the deterministic phase.

A typical problem will have state variables of many kinds: costs of manufacture, prices charged by competitors, the failure rate of the product, etc. We select them by asking the question, "If you had a crystal ball, what numerical questions would you ask it about the outcome in order to specify your profit measure?" At the same time that we select these variables we should assign both nominal values for them and the range over which they might vary for future reference.

Next we establish how the state variables are related to each other and to the measure of performance. We construct, in essence, a profit function that shows how profit is related to the factors that underlie the decision. The construction of this profit function requires considerable judgment to avoid the twin difficulties of excessive complexity and unreal simplicity.

If the results of the decision extend over a long time period, it will be necessary to have the decision maker specify his time preference for profit. We must

ask, "How does profit received in the future compare in value to profit received today?" or an equivalent question. In cases in which we can assume a perfect financial environment the present value of future profit at some rate of interest will be the answer. In many large decision problems, however, the nature of the undertaking has an effect on the basic financial structure of the enterprise. In these cases a much more realistic modeling of the time preference for profit is necessary.

Now that we have completed the steps in the deterministic phase we have a deterministic model of the decision problem. We next perform two closely related analyses. We perform them by setting the state variables to their nominal values and then sweeping each through its range of values, individually and jointly, as judgment dictates. Throughout this process we observe which alternative would be best and how much value would be associated with each alternative. We often observe that regardless of the values the state variables take on in their ranges one alternative is always superior to another, a condition we describe by saying that the first alternative dominates the second. The principle of dominance may often permit a major reduction in the number of alternatives that need be considered.

As a result of this procedure we have performed a sensitivity analysis on the state variables. We know how much a 10 percent change in one of the variables will affect profit, hence the optimum alternative. Similarly, we know how changes in state variables may interact to affect the decision. This sensitivity analysis shows us where uncertainty is important. We identify those state variables to which the outcome is sensitive as "crucial" state variables. Determining how uncertainties in the crucial state variable influence the decision is the concern of the probabilistic phase of the decision analysis.

2.1.2. *Probabilistic Phase*

The probabilistic phase begins by encoding uncertainties on each of the crucial state variables; that is, gathering priors on them. A subset of the crucial state variables will usually be independent—for these only a single probability distribution is necessary. The remainder will have to be treated by collecting conditional as well as marginal distributions. We have more to say on this process later.

The next step is to find the uncertainty in profit for each alternative implied by the functional relationship of profit to the crucial state variables and the probability distribution on those crucial state variables for the alternative. We call this derived probability distribution of profit the profit lottery of the alternative. In a few cases the profit lottery can be derived analytically and in many by numerical analysis procedures. In any case it may be approximated by a Monte Carlo simulation. Regardless of the procedure used, the result is a probability distribution on profit (or perhaps on discounted profit) for each of the alternatives that remain in the problem.

Now we must consider how to choose between two alternatives with different profit lotteries. In one case the choice is easy. Suppose that we plot the profit lottery for each alternative in complementary cumulative form; that is, plot the

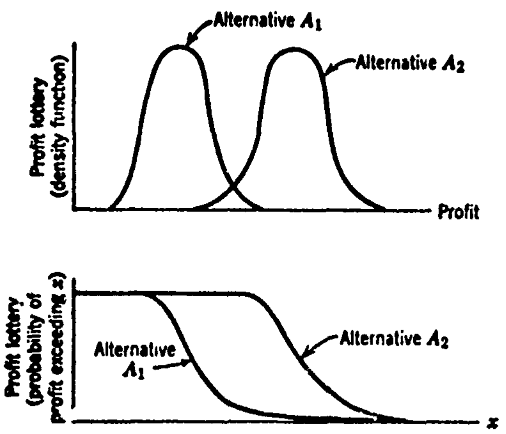


Figure 1. Stochastic dominance.

probability of profit exceeding x for any given x. Suppose further, as shown in Figure 1, that the complementary cumulative for alternative A_2 always lies above that for alternative A_1. This means that for any number x there is a higher probability of profit exceeding that number with alternative A_2 than with alternative A_1. In this case we would prefer alternative A_2 to alternative A_1, provided only that we liked more profit better than less profit. We describe this situation by saying that the profit from alternative A_2 is stochastically greater than the profit from alternative A_1 or equivalently by saying that alternative A_2 stochastically dominates alternative A_1. Stochastic dominance is a concept that appeals intuitively to management; it applies in a surprising number of cases.

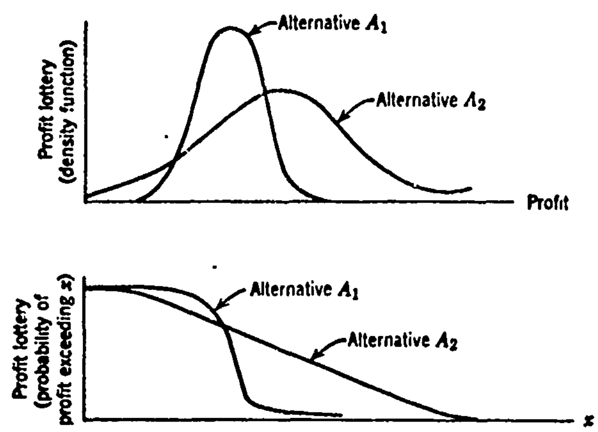


Figure 2. Lack of stochastic dominance.

Figure 2, however, illustrates a case in which stochastic dominance does not apply. When faced with a situation like this, we must either abandon formal methods and leave the selection of the best alternative to judgment or delve into the measurement of risk preference. If we choose to measure risk preference, we begin the second step of the probabilistic phase. We must construct a utility function for the decision maker that will tell us whether or not, for example, he would prefer a certain 4 million dollars profit to equal chances of earning zero or 10 million dollars. Although these questions are quite foreign to management, they are being asked increasingly often with promising results. Of course, when risk preference is established in the form of a utility function, the best alternative is the one whose profit lottery has the highest utility.

2.1.3. Post-Mortem Phase

The post-mortem phase of the procedure is composed entirely of analysis. This phase begins when the best alternative has been selected as the result of the probabilistic phase. Here we use the concepts of the clairvoyant lottery to establish a dollar value of eliminating uncertainty in each of the state variables individually and jointly. Being able to show the impact of uncertainties on profit is one of the most important features of decision analysis. It leads directly to the next step of the post-mortem, which is finding the most economical information-gathering program, if, in fact, it would be profitable to gather more information. The information-gathering program may be physical research, a marketing survey, or the hiring of a consultant. Perhaps in no other area of its operations is an enterprise in such need of substantiating analysis as it is in the justification of information-gathering programs.

Of course, once the information-gathering scheme, if any, is completed, its information modifies the probability distributions on the crucial state variables and consequently affects the decision. Indeed, if the information-gathering program were not expected to modify the probability distributions on the crucial state variables it would not be conducted. We then repeat the probabilistic phase by using the new probability distributions to find the profit lotteries and then enter the post-mortem phase once more to determine whether further information gathering is worthwhile. Thus the decision analysis is a vital structure that lets us compare at any time the values of such alternatives as acting, postponing action and buying information, or refusing to consider the problem further. We must remember that the analysis is always based on the current state of knowledge. Overnight there can arrive a piece of information that changes the nature of the conclusions entirely. Of course, having captured the basic structure of the problem, we are in an excellent position to incorporate any such information.

Finally, as the result of the analysis the decision maker embarks on a course of action. At this point he may be interested in the behavior of several of the state variables for planning purposes; for example, having decided to introduce a new product, he may want to examine the probability distributions for its sales in future years to make subsidiary decisions on distribution facilities or

on the size of the sales force. The decision-analysis model readily provides such planning information.

2.2. The Advantages of Decision Analysis

Decision analysis has many advantages, of which we have described just a few, such as its comprehensiveness and vitality as a model of the decision and its ability to place a dollar value on uncertainty. We should point out further that the procedure is relevant to both one of a kind and repetitive decisions. Decision analysis offers the operations research profession the opportunity to extend its scope beyond its traditional primary concern with repetitively verifiable operations.

One of the most important advantages of decision analysis lies in the way it encourages meaningful communication among the members of the enterprise because it provides a common language in which to discuss decision problems. Thus engineers and marketing planners with quite different jargons can appreciated one another's contributions to a decision. Both can use the decision-analysis language to convey their feelings to management quickly and effectively.

A phenomenon that seems to be the result of the decision-analysis language is the successive structuring of staff groups to provide reports that are useful in decision-analysis terms. Thus, if the decision problem being analyzed starts in an engineering group, that group ultimately seeks inputs from marketing, product planning, the legal staff, and so on, that are compatible with the probabilistic analysis. Soon these groups begin to think in probabilistic terms and to emphasize probabilistic thinking in their reports. The process seems irreversible in that once the staff of an organization becomes comfortable in dealing with probabilistic phenomena they are never again satisfied with deterministic or expected value approaches to problems. Thus the existence of decision-analysis concepts as a language for communication may be its most important advantage.

2.3. The Hierarchy of Decision Analysis

It is informative to place decision analysis in the hierarchy of techniques that have been developed to treat decision problems. We see that a decision analysis requires two supporting activities. One is a lower order activity that we call alternative evaluation; the second, a higher order activity that we call goal setting. Performing a decision analysis requires evaluating alternatives according to the goals that have been set for the decision. The practitioners of operations research are quite experienced in alternative evaluation in both industrial and military contexts. In fact, in spite of the lip service paid to objective functions, only rare operations researchers have had the scope necessary to consider the goal-setting problems.

All mankind seems inexpert at goal setting, although it is the most important problem we face. Perhaps the role of decision analysis is to allow the discussion of decisions to be carried on at a level that shows the explicit need for goals or criteria for selection of the best alternative. We need to make goals explicit only

if the decision maker is going to delegate the making of the decision or if he is unsure of his ability to be consistent in selecting the best alternative. We shall not comment on whether there is a trend toward more or less delegation of decision making. However, it is becoming clear to those with decision-making responsibilities that the increasing complexity of the operations under their control requires correspondingly more formal approaches to the problem of organizing the information that bears on a decision if inconsistent decisions are to be avoided.

The history of the analysis of the procurement of military weapons systems points this out. Recent years have shown the progression of procurement thinking from effectiveness to cost effectiveness. In this respect the military authorities have been able to catch up in their decision-making apparatus to what industry had been doing in its simpler problems for years. Other agencies of government are now in the process of making the same transition. Now all must move on to the inclusion of uncertainty, to the establishment of goals that are reflected in risk and time preferences.

These developments are now on the horizon and in some cases in sight; for example, although we have tended to think of the utility theory as an academic pursuit, one of our major companies was recently faced with the question, "Is 10 million dollars of profit sufficient to incur one chance in 1 million of losing 1 billion dollars?" Although the loss is staggering, it is realistic for the company concerned. Should such a large company be risk-indifferent and make decisions on an expected value basis? Are stockholders responsible for diversifying their risk externally to the company or should the company be risk-averting on their behalf? For the first time the company faced these questions in a formal way rather than deciding the particular question on its own merits and this we must regard as a step forward.

Decision analysis has had its critics, of course. One said, "In the final analysis, aren't decisions politically based?" The best answer to that came from a high official in the executive branch of our government who said, "The better the logical basis for a decision, the more difficult it is for extraneous political factors to hold sway." It may be discouraging in the short run to see logic overridden by the tactical situation, but one must expect to lose battles to win the war.

Another criticism is, "If this is such a good idea, why haven't I heard of it before?" One very practical reason is that the operations we conduct in the course of a decision analysis would be expensive to carry out without using computers. To this extent decision analysis is a product of our technology. There are other answers, however. One is that the idea of probability as a state of mind and not of things is only now regaining its proper place in the world of thought. The opposing heresy lay heavy on the race for the better part of a century. We should note that most of the operations research performed in World War II required mathematical and probabilistic concepts that were readily available to Napoleon. One wonders about how the introduction of formal methods for decision making at that time might have affected the course of history.

3. THE PRINCIPLES OF THE DECISION ANALYST

Next we turn to the principles of the decision analyst, the professional who embarks on preparing a decision analysis. His first principle is to identify and isolate the components of the decision—the uncertainty, risk aversion, time preference, and problem structure. Often arguments over which is the best decision arise because the participants do not realize that they are arguing on different grounds. Thus it is possible for A to think that a certain alternative is riskier than it is in B's opinion, either because A assigns different probabilities to the outcomes than B but both are equally risk-averting, or because A and B assign the same probabilities to the outcomes but differ in their risk aversion. If we are to make progress in resolving the argument, we must identify the nature of the difficulty and bring it into the open. Similar clarifications may be made in the areas of time preference or in the measurement of the value of outcomes.

One aid in reducing the problem to its fundamental components is restricting the vocabulary that can be used in discussing the problem. Thus we carry on the discussion in terms of events, random variables, probabilities, density functions, expectations, outcomes, and alternatives. We do not allow fuzzy thinking about the nature of these terms. Thus "The density function of the probability" and "The confidence in the probability estimate" must be nipped in the bud. We speak of "assigning," not "estimating," the probabilities of events and think of this assignment as based on our "state of information." These conventions eliminate statements like the one recently made on a TV panel of doctors who were discussing the right of a patient to participate in decision making on his treatment. One doctor asserted that the patient should be told of "some kind of a chance of a likelihood of a bad result." I am sure that the doctor was a victim of the pressures of the program and would agree with us that telling the patient the probability the doctor would assign to a bad result would be preferable.

One principle that is vital to the decision analyst is professional detachment in selecting alternatives. The analyst must not become involved in the heated political controversies that often surround decisions except to reduce them to a common basis. He must demonstrate his willingness to change the recommended alternative in the face of new information if he is to earn the respect of all concerned. This professional detachment may, in fact, be the analyst's single most valuable characteristic. Logic is often severely strained when we are personally involved.

The detachment of the analyst has another positive benefit. As an observer he may be able to suggest alternatives that may have escaped those who are intimately involved with the problem. He may suggest delaying action, buying insurance, or performing a test, depending on the nature of the decision. Of course, the comprehensive knowledge of the properties of the existing alternatives that the decision analyst must gain is a major aid in formulating new alternatives.

Since it is a rare decision that does not imply other present and future decisions, the decision analyst must establish a scope for the analysis that is

broad enough to provide meaningful answers but not broad enough to impose impractical computational requirements. Perhaps the fundamental question in establishing scope is how much to spend on decision analysis. Because the approach could be applied both to selecting a meal from a restaurant menu and to allocating the federal budget, the analyst needs some guidelines to determine when the analysis is worthwhile.

The question of how much decision analysis is an economic problem susceptible to a simpler decision analysis, but rather than pursue that road let us pose an arbitrary and reasonable but indefensible rule of thumb: spend at least 1 percent of the resources to be allocated on the question of how they should be allocated. Thus, if we were going to buy a 2000-dollar automobile, the rule indicates a 20-dollar analysis, whereas for a 20,000-dollar house it would specify a 200-dollar analysis. A 1-million-dollar decision would justify 10,000 dollars' worth of analysis or, let us say, about three man-months. The initial reaction to this guideline has been that it is conservative in the sense of not spending much on analysis; yet, when we apply it to many decisions now made by business and government, the reaction is that the actual expenditures on analysis are only one-tenth or one-hundredth as large as the rule would prescribe. Of course, we can all construct situations in which a much smaller or larger expenditure than given by the rule would be appropriate, and each organization can set its own rule, perhaps making the amount spent on analysis nonlinear in the resources to be allocated. Nevertheless, the 1 percent figure has served well to illustrate where decision analysis can be expected to have the highest payoff.

The professional nature of the decision analyst becomes apparent when he balances realism in the various parts of the decision-analysis model. Here he can be guided only by what used to be called engineering judgment. One principle he should follow is to avoid sophistication in any part of the problem when that sophistication would not affect the result. We can describe this informally by saying that he should strive for a constant "wince" level as he surveys all parts of the analysis. One indication that he has achieved this state is that he would be torn among many possibilities for improvement if we allowed him to devote more time and resources to the decision model.

4. THE ENCODING OF SUBJECTIVE INFORMATION

One unique feature of decision analysis is the encoding of subjective information, both in the form of risk aversion and in the assignment of probabilities.

4.1. Risk Aversion and Time Preference

Since we are dealing in most cases with enterprises rather than individuals, the appropriate risk aversion and time preference should be that of the enterprise. The problem of establishing such norms is beyond our present scope. It is easy, however, to demonstrate to managers, or to anyone else for that matter, that the phenomenon of risk aversion exists and that it varies widely from individual to individual. One question useful in doing this is, "How much would you have to be paid to call a coin, double or nothing, for next year's

salary?" Regardless of the salary level of the individuals involved, this is a provocative question. We point out that only a rare individual would play such a game for a payment of zero and that virtually everyone would play for a payment equal to next year's salary, since then there would be nothing to lose. Thereafter we are merely haggling over the price. Payments in the range of 60 percent to 99 percent of next year's salary seem to satisfy the vast majority of professional individuals.

The steps required to go from a realization of personal risk aversion and time preference to corporate counterparts and finally to a reward system for managers that will encourage them to make decisions consistent with corporate risk aversion and time preference remain a fascinating area of research.

4.2. Encoding of Uncertainty

When we begin the probabilistic phase of the decision analysis, we face the problem of encoding the uncertainty in each of the crucial state variables. We shall want to have the prior probability distributions assigned by the people within the enterprise who are most knowledgeable about each state variable. Thus the priors on engineering variables will typically be assigned by the engineering department; on marketing variables, by the marketing department, and so on. However, since we are in each case attempting to encode a probability distribution that reflects a state of mind and since most individuals have real difficulty in thinking about uncertainty, the method we use to extract the priors is extremely important. As people participate in the prior-gathering process, their attitudes are indicated successively by, "This is ridiculous," "It can't be done," "I have told you what you want to know but it doesn't mean anything," "Yes, it seems to reflect the way I feel," and "Why doesn't everybody do this?" In gathering the information we must be careful to overcome the defenses the individual develops as a result of being asked for estimates that are often a combination of targets, wishful thinking, and expectations. The biggest difficulty is in conveying to the man that you are interested in his state of knowledge and not in measuring him or setting a goal for him.

If the subject has some experience with probability, he often attempts to make all his priors look like normal distributions, a characteristic we may designate as "bellshaped" thinking. Although normal distributions are appropriate priors in some circumstances, we must avoid making them a foregone conclusion.

Experience has shown certain procedures to be effective in this almost psychoanalytic process of prior measurement. The first procedure is to make the measurement in a private interview to eliminate group pressure and to overcome the vague notions that most people exhibit about matters probabilistic. Sending around forms on which the subjects are supposed to draw their priors has been worse than useless, unless the subjects were already experienced in decision analysis.

Next we ask questions of the form, "What are the chances that x will exceed 10," because people seem much more comfortable in assigning probabilities to events than they are in sketching a density function. As these questions are

asked, we skip around, asking the probability that x will be "greater than 50, less than 10, greater than 30," often asking the same question again later in the interview. The replies are recorded out of the view of the subject in order to frustrate any attempt at forced consistency on his part. As the interview proceeds, the subject often considers the questions with greater and greater care, so that his answers toward the end of the interview may represent his feelings much better than his initial answers. We can change the form of the questions by asking the subject to divide the domain of the random variable into n mutually exclusive regions with equal probability. (Of course, we would never put the question to him that way.) We can use the answers to all these questions to draw the complementary cumulative distribution for the variable, a form of representation that seems easiest to convey to people without formal probabilistic training.

The result of this interview is a prior that the subject is willing to live with, regardless of whether we are going to use it to govern a lottery on who buys coffee or on the disposal of his life savings. We can test it by comparing the prior with known probabilistic mechanisms; for example, if he says that a is the median of the distribution of x, then he should be indifferent about whether we pay him one hundred dollars if x exceeds a or if he can call the toss of a coin correctly. If he is not indifferent, then we must require him to change a until he is. The end result of such questions is to produce a prior that the subject is not tempted to change in any way, and we have thus achieved our final goal. The prior-gathering process is not cheap, but we perform it only on the crucial state variables.

In cases in which the interview procedure is not appropriate, the analyst can often obtain a satisfactory prior by drawing one himself and then letting the subject change it until the subject is satisfied. This technique may also be useful as an educational device in preparation for the interview.

If two or more variables are dependent, we must gather priors on conditional as well as marginal distributions. The procedure is generally the same but somewhat more involved. However, we have the benefit of being able to apply some checks on our results. Thus, if we have two dependent variables x and y, we can obtain the joint distribution by measuring the prior on x and the conditional on y, given x, or, alternatively, by measuring the prior on y and the conditional on x, given y. If we follow both routes, we have a consistency check on the joint distribution. Since the treating of joint variables is a source of expense, we should formulate the problem to avoid them whenever possible.

To illustrate the nature of prior gathering we present the example shown in Figure 3. The decision in a major problem was thought to depend primarily on the average lifetime of a new material. Since the material had never been made and test results would not be available until three years after the decision was required, it was necessary to encode the knowledge the company now had concerning the life of the material. This knowledge resided in three professional metallurgists who were experts in that field of technology. These men were interviewed separately according to the principles we have described. They produced the points labeled "Subjects 1, 2, and 3" in Figure 3. These results have several interesting features. We note, for example, that for $t = 17$ Subject

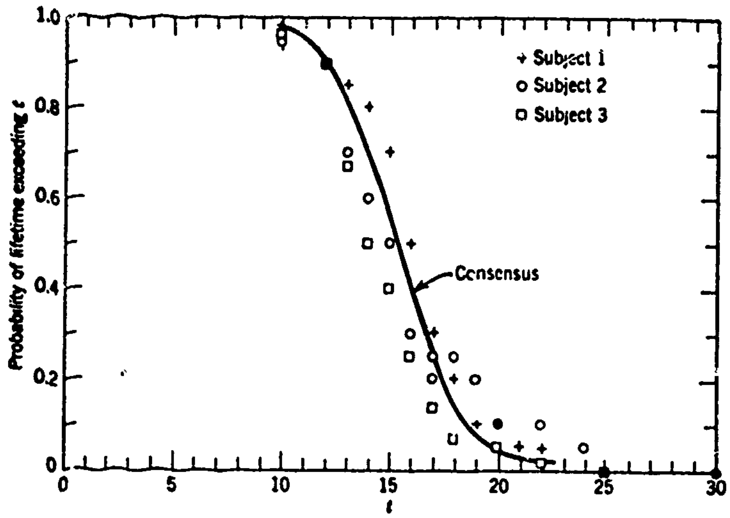


Figure 3. Priors on lifetime of material.

2 assigned probability 0.2 and 0.25 at various points in the interview. On the whole, however, the subjects were remarkably consistent in their assignments. We observe that Subject 3 was more pessimistic than Subject 1.

At the conclusion of the three interviews the three subjects were brought together and shown the results. At this point a vigorous discussion took place. Subjects 1 and 3, in particular brought forth information of which the other two members of the group were unaware. As the result of this information exchange, the three group members drew the consensus curve—each subject said that this curve represented the state of information about the material life at the end of the meeting.

It has been suggested that the proper way to reconcile divergent priors is to assign weights to each, multiply, and add, but this experiment is convincing evidence that any such mechanistic procedure misses the point. Divergent priors are an excellent indicator of divergent states of information. The experience just described not only produced the company's present encoding of uncertainty about the lifetime of the material but at the same time encouraged the exchange of information within the group.

5. A DECISION-ANALYSIS EXAMPLE

To illustrate the flavor of application let us consider a recent decision analysis in the area of product introduction. Although the problem was really from another industry, let us suppose that it was concerned with the development and production of a new type of aircraft. There were two major alternatives: to develop and sell a new aircraft (A_2) or to continue manufacturing and selling

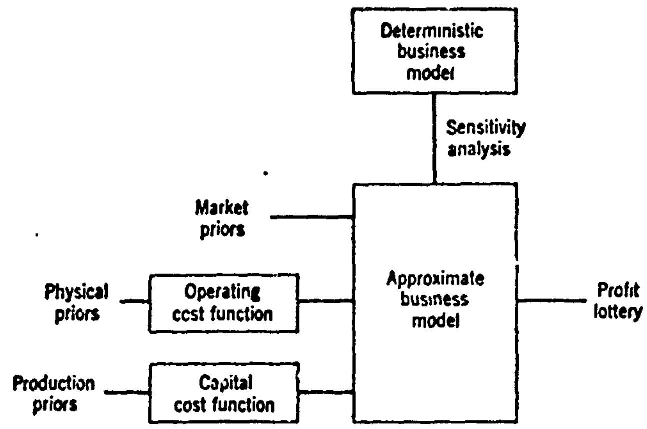


Figure 4. Decision analysis for new product introduction.

the present product (A_1). The decision was to be based on the present value of future expected profits at a discounting rate of 10 percent per year. Initially, the decision was supposed to rest on the lifetime of the material for which we obtained the priors in Figure 3; however, a complete decision analysis was desired. Since several hundred million dollars in present value of profit were at stake, the decision analysis was well justified.

The general scheme of the analysis appears in Figure 4. The first step was to construct a model of the business, a model that was primarily a model of the market. The profit associated with each alternative was described in terms of the price of the product, its operating capital costs, the behavior of its competitors, and the national characteristics of customers. The actual profit and discounted profit were computed over a 22-year time period. A suspicion grew that this model did not adequately capture the regional nature of demand. Consequently a new model was constructed that included the market characteristics, region by region and customer by customer. Moving to the more detailed basis affected the predictions so much that the additional refinement was clearly justified. Other attempts at refinement, however, did not affect the results sufficiently to justify a still more refined model. Now, the sensitivity analysis was performed to determine the crucial state variables, which turned out to be the operating cost, capital cost, and a few market parameters. Because of the complexity of the original business model, an approximate business model essentially quadratic in form was constructed to show how profit depended on these crucial state variables in the domain of interest. The coefficients of the approximate business model were established by runs on the complete business model.

The market priors were directly assigned with little trouble. However, because the operating and capital costs were the two most important variables

in the problem, these priors were assigned according to a more detailed procedure. First, the operating cost was related to various physical features of the design by the engineering department. This relationship was called the operating-cost function. One of the many input physical variables was the average lifetime of the material whose priors appear in Figure 3. All but two of the 12 physical input variables were independent. The priors on the whole set of input variables were gathered and used with the operating-cost function in a Monte Carlo simulation that produced a prior for the operating cost of the product.

The capital-cost function was again developed by engineering but was much simpler in form. The input certainties were the production costs for various parts of the product. Again, a Monte Carlo analysis produced a prior on capital cost..

Once we had established priors on all inputs to the approximate business model, we could determine the profit lottery for each alternative, in this case by using numerical analysis.

The present-value profit lotteries for the two alternatives looked very much like those shown in Figure 1. The new product alternative A_2 stochastically dominated the alternative A_1 of continuing to manufacture the present product. The result showed two interesting facets of the problem. First, it had been expected that the profit lottery for the new product alternative would be considerably broader than it was for the old product. The image was that of a profitable and risky new venture compared with a less profitable but less risky standard venture. In fact, the results showed that the uncertainties in profit were about the same for both alternatives, thus showing how initial concepts may be misleading.

The second interesting facet was that the average lifetime of the material whose priors appear in Figure 3 was actually of little consequence in the decision. It was true enough that profits were critically dependent on this lifetime if the design were fixed, but if the design were left flexible to accommodate to different average material lifetimes profits would be little affected. Furthermore, leaving the design flexible was not an expensive alternative; therefore another initial conception had to be modified.

However, the problem did not yield so easily. Figure 5 shows the present value of profits through each number of years t for each alternative. Note that if we ignore returns beyond year 7 the new product has a higher present value but that if we consider returns over the entire 22-year period the relationship reverses, as we have already noted. When management saw these results, they were considerably disturbed. The division in question had been under heavy pressure to show a profit in the near future—alternative A_2 would not meet that requirement. Thus the question of time preference that had been quickly passed off as one of present value at 10 percent per year became the central issue in the decision. The question was whether the division was interested in the quick kill or the long pull. At last report the division was still trying to convince the company to extend its profit horizon.

This problem clearly illustrates the use of decision analysis in clarifying the

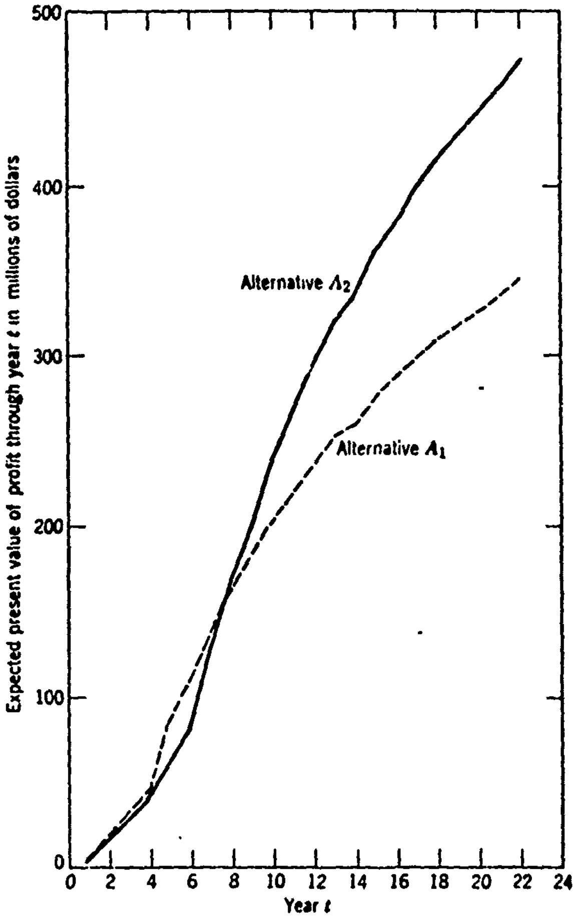


Figure 5. Expected present value of profit.

issues surrounding a decision. A decision that might have been made on the basis of a material lifetime was shown to depend more fundamentally on the question of time preference for profit. The nine man-months of effort devoted to this analysis were considered well spent by the company. The review committee for the decision commented, "We have never had such a realistic analysis of a new business venture before." The company is now interested in instituting decision-analysis procedures at several organizational levels.

6. CONCLUSION

Decision analysis offers operations research a second chance at top management. By foregoing statistical reproducibility we can begin to analyze the one-of-a-kind problems that managers have previously had to handle without assistance. Experience indicates that the higher up the chain of management we progress the more readily the concepts we have outlined are accepted. A typical reaction is, "I have been doing this all along, but now I see how to reduce my ideas to numbers."

Decision analysis is no more than a procedure for applying logic. The ultimate limitation to its applicability lies not in its ability to cope with problems but in man's desire to be logical.

ANALYSE DES DECISIONS: THEORIE APPLIQUEE DES DECISIONS

Reproduced from best available copy.

RÉSUMÉ

Au cours de ces dernières années, la théorie de décision a été de plus en plus acceptée en tant que cadre conceptuel pour la prise de décision. Cependant, cette théorie a surtout affecté les statisticiens plutôt que les personnes qui en ont le plus besoin: les responsables de décisions. Cette étude décrit un procédé qui permet de replacer des problèmes de décision réels dans la structure de la théorie de décision. Le procédé d'analyse de décision englobe chaque étape, du mesurage des choix de risques et des jugements portant sur des facteurs critiques par l'établissement de structures des facteurs relatifs à la technique, au marché, à la rivalité commerciale et à l'environnement, jusqu'au mesurage des préférences subjectives et de la valeur de la prédiction. L'analyse de décision met en perspective les nombreux instruments de simulation, d'analyse numérique, et de transformations de probabilités qui deviennent de plus en plus commodes depuis le développement des systèmes d'ordinateurs électroniques dont les différentes "stations" dépendent d'une "centrale" unique.

Le procédé est appliqué à un problème de décision réelle qui s'étend sur des dizaines d'annés et dont la valeur actuelle est de plusieurs centaines de millions de dollars. Cette étude analyse le problème de la détermination des dépenses consacrées à l'analyse de décisions. L'une des plus importantes propriétés de ce procédé tient au nombre des bénéfices auxiliaires créés au cours de l'élaboration de ce genre d'étude. L'expérience montre que ces bénéfices peuvent excéder en valeur le coût des dépenses consacrées a l'élaboration de la décision.

Decision Analysis Practice: Examples and Insights

J. E. Matheson

Introduction

Decision analysis is a discipline that merges the logical foundations of statistical decision theory with the capabilities of modeling and solving complex problems developed in the fields of systems analysis and operations research.[1,2] Statistical decision theory forms both a logical structure for describing the uncertainties, values, and preferences that are relevant to a decision and a set of mathematical techniques for treating problems in which uncertainty is a factor. The fields of systems analysis and operations research provide the methodology for applying abstract models to complex, real-world situations. Together these foundations yield the new discipline of decision analysis. Using the decision to be made as the focal point of the analysis, the analyst tailors his modeling and information gathering efforts to the specific decision. In this paper I will describe the professional practice of decision analysis and will present several applications of it that are familiar to me.

Bounding the Decision Problem

In approaching a problem, the decision analyst's first responsibility is to define clearly the decision to be made. Since most, if not all, decision problems are subordinate to some higher-level system, it is vitally important to bound the decision problem; that is, to establish who has the responsibility for making the decision, to determine what resources are to be allocated, and to set out which values and preferences are to be delegated explicitly by the higher-level system and which ones are to be specified by the decision-maker. For example, if the decision calls for allocating funds for new capital investments, the analyst might

Preceding page blank

decide to use interest rates derived from a higher-level financial system and to use present worth of profits as the measure of value. However, if the decision calls for securing financing, considering the characteristics of each method of financing might well be within the bounds of the problem. Many times a problem is "difficult" because of the way in which the boundaries of the problem have been specified. In many cases, the analyst can transcend such difficulties by changing the specification of the bounds.

Establishing the Extensiveness of the Analysis

The extent of the analysis that should be applied to any decision problem depends on the value of the resources that are at stake and the likelihood that the analysis will improve the outcome of the action taken through the selection of a "better" decision. In fact, establishing the economic value of the analysis is a decision analysis in itself.[3] However, generally the amount of resources being allocated to the analysis is too small to justify such formal treatment.

In practice, an attempt is usually made to carry out a simplified analysis of the entire decision problem. Techniques such as sensitivity analysis and determination of the value of perfect (and sometimes imperfect) information indicate where the model should be refined and the kind of new information that should be gathered. In many cases, the analysis effort goes through three stages. The first is the pilot stage, in which the conceptual structure of the analysis is created and tested, while many of the detailed features of the problem are suppressed. During the next stage, the prototype stage, a more detailed analysis is carried out in an attempt to capture all of the relevant features of the

problem. This stage is likely to involve the development of large computer models. The final stage is the production analysis, in which all aspects of the problem are critically reviewed and a decision is recommended. The decision may, of course, be a decision to gather more information and incorporate it into the analysis before making the final decision.

Relationship Between the Analyst and His Client

The decision analyst usually serves a decision-maker or a decision-making body that I will call the client. The decision analyst is expert only in his discipline, while the client holds the resources, and knows the information, the values, and the preference that form the decision problem. If the analyst is to conduct an unbiased analysis, he must be careful to encode only his client's information and avoid biasing his analysis by inserting his own opinions.

To allow the analyst to maintain this division, the client must clearly designate who will be responsible for supplying various kinds of information, values, and preferences. In complex problems, much of the information is encoded in the structure of the model itself. Building and verifying the decision analysis model requires an interaction between the analyst and client that is perhaps the most difficult and challenging part of the task.

Examples

In the rest of this paper, I will present three examples of applications for purposes of illustrating the practice of decision analysis. The first is a "typical application" to a new product development decision. The second is the result of decision analysis research on space program

planning. The last is a large-scale application to planning for an electrical power system.

New Product Development

A major manufacturing research company had developed two compounds for a particular market. Compound A was developed and tested to the point where it was well beyond the research stage and one alternative was to develop it into the final product. Another alternative was to develop compound B, which was still in the research stage, but was thought to be more potent than compound A. A third alternative was to abandon the whole effort.

It was thought that the development of the new product would be lengthy and expensive and that the potential market was very uncertain. Since this was a new marketing area for the client, he engaged an outside expert to carry out a market survey for use as one of the informational inputs to the decision analysis.

The analysis followed quite closely the decision analysis cycle displayed in Figure 1. The deterministic phase was begun by laying out the decision tree shown in Figure 2. The first decision was whether to develop compound A or compound B (or both) into a final product. This development determined the production cost of the compound and the concentration of it that would be required in the final product. After this determination was complete the choice of whether to market or abandon the product could be made. There were still uncertainties about the size and growth rate of the market and the action of competitors. These additional facets of the problem were represented in the structural model shown in Figure 3.

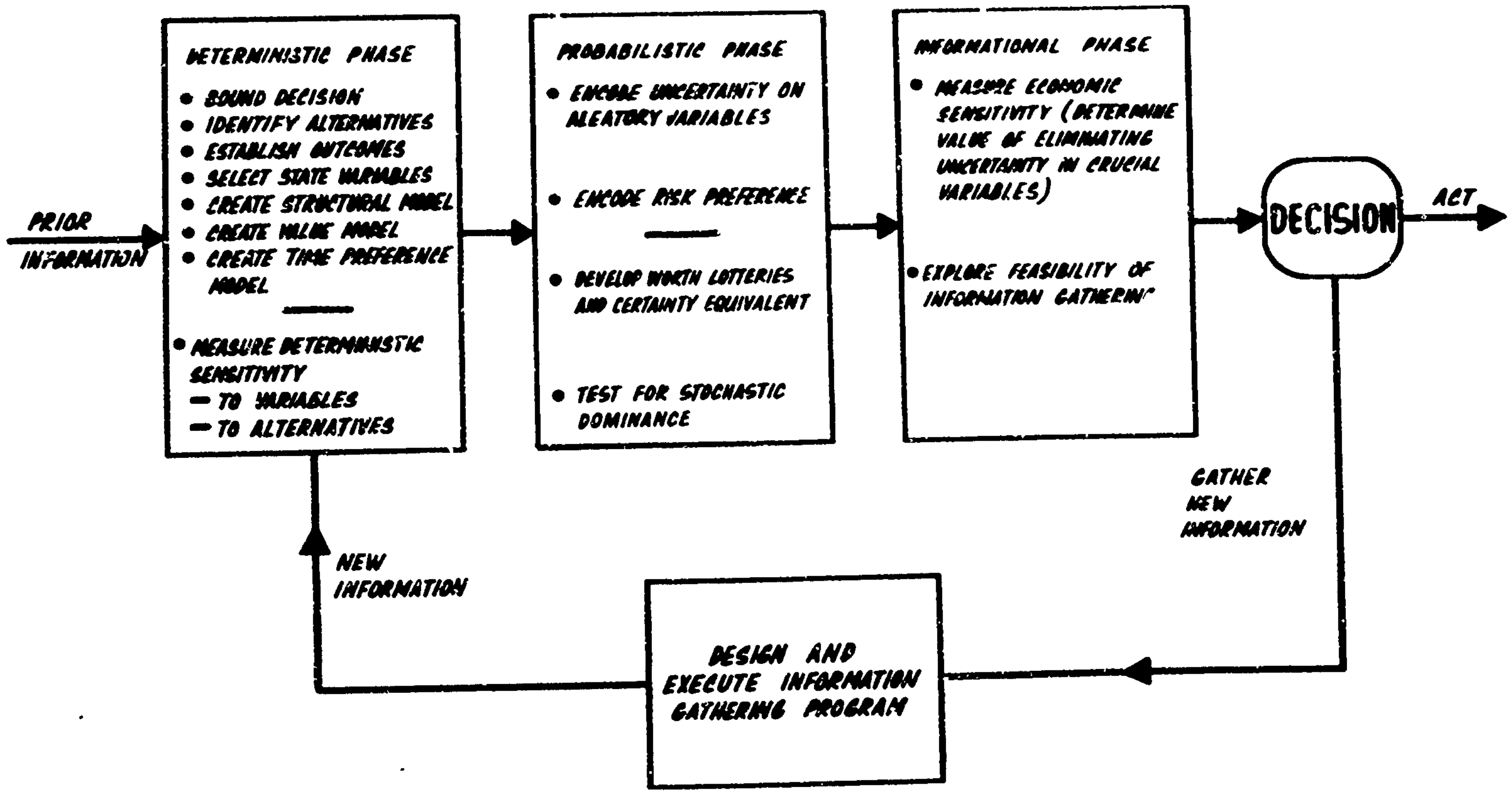


FIG 1 THE DECISION ANALYSIS CYCLE

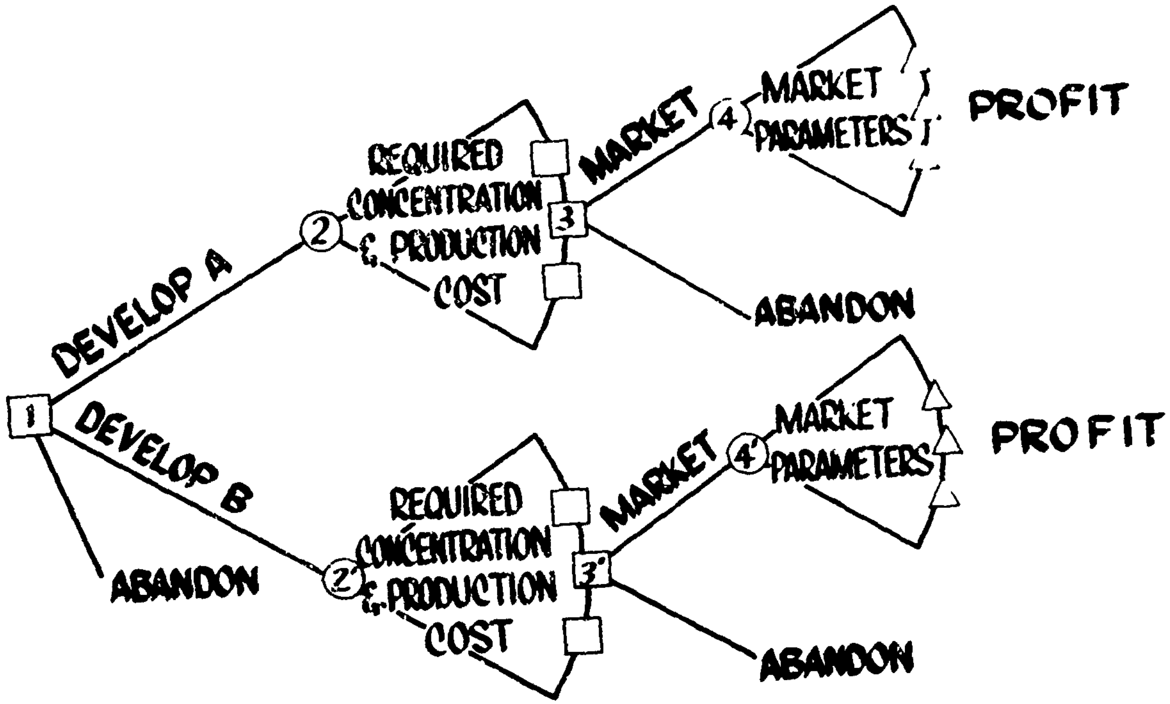


FIG 2 SIMPLIFIED DECISION TREE FOR THE NEW PRODUCT INTRODUCTION

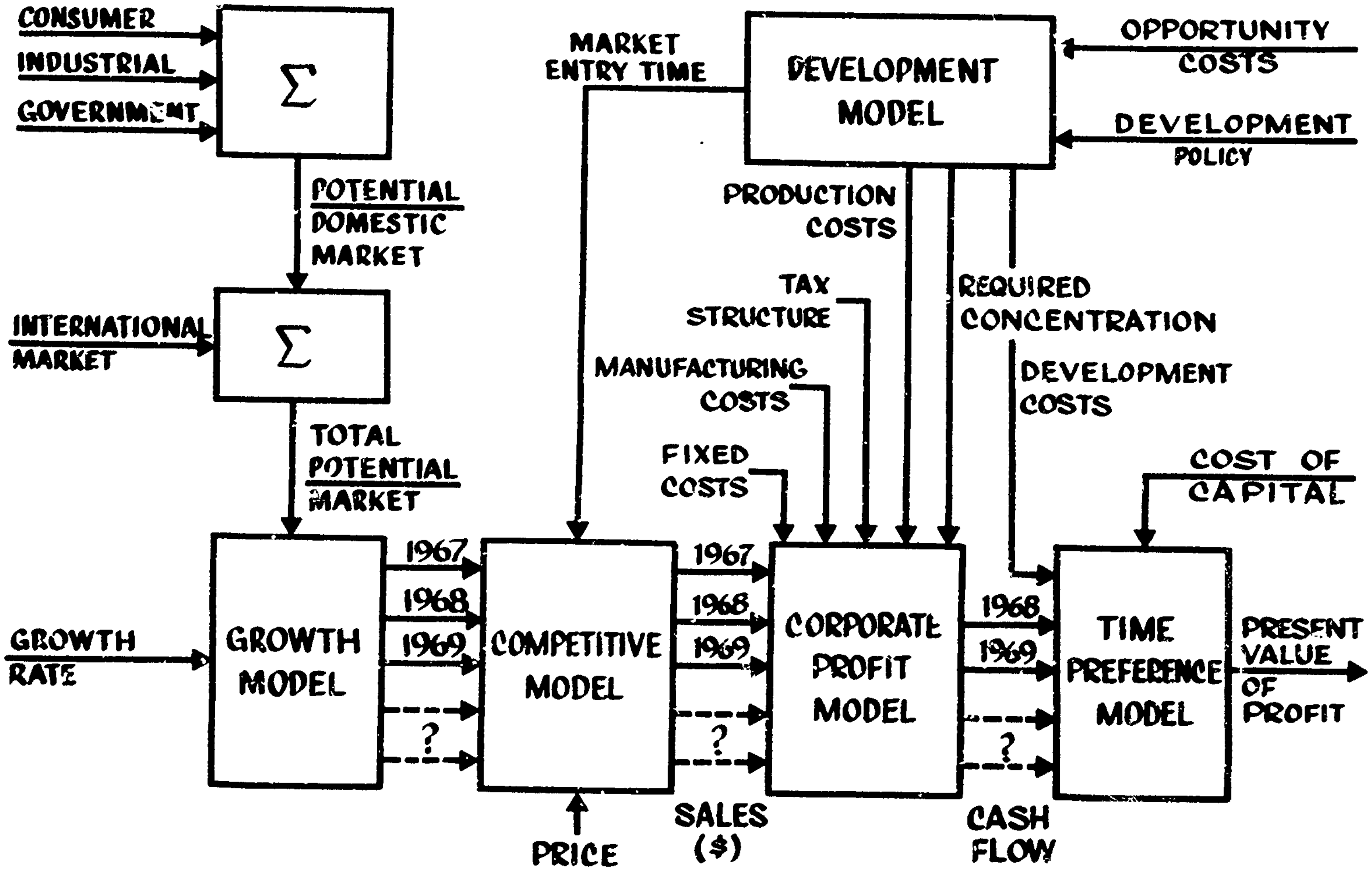


FIG 3 STRUCTURAL MODEL FOR THE NEW PRODUCT INTRODUCTION

Many of the variables in the problem were subjected to sensitivity analysis. The most sensitive variable, international market size, produced changes of 16 million dollars in the present value of profit. Five variables were selected as aleatory variables--variables whose uncertainty is encoded in terms of probability distributions--for the probabilistic phase.

In the probabilistic phase, the simplified decision tree (Figure 2) was developed into a detailed decision tree, assigning actual conditional probabilities to the aleatory variables represented in the structure of the tree. At the tips of this tree, expected profits were assigned by a Monte Carlo simulation of the structural model of Figure 3, which contained the remaining aleatory variables. The decision tree was then evaluated on an expected value basis. The amount of corporate resources to be devoted to this product were small enough so that no significant risk aversion was desirable.

The result of the probabilistic phase was that the profit lottery for development of compound B stochastically dominated that of compound A. However, the profit lottery for the development of compound B, with the cumulative probability distribution shown in Figure 4, had negative expected present value, so the best decision was to abandon the effort.

In the informational phase, the expected value of perfect information (economic sensitivity) was computed on several important aleatory variables. The highest economic sensitivity of $1,415,000 was exhibited by the international market size. The international market size showed such high economic sensitivity because the new information might reveal a very large international market for the product, making it profitable

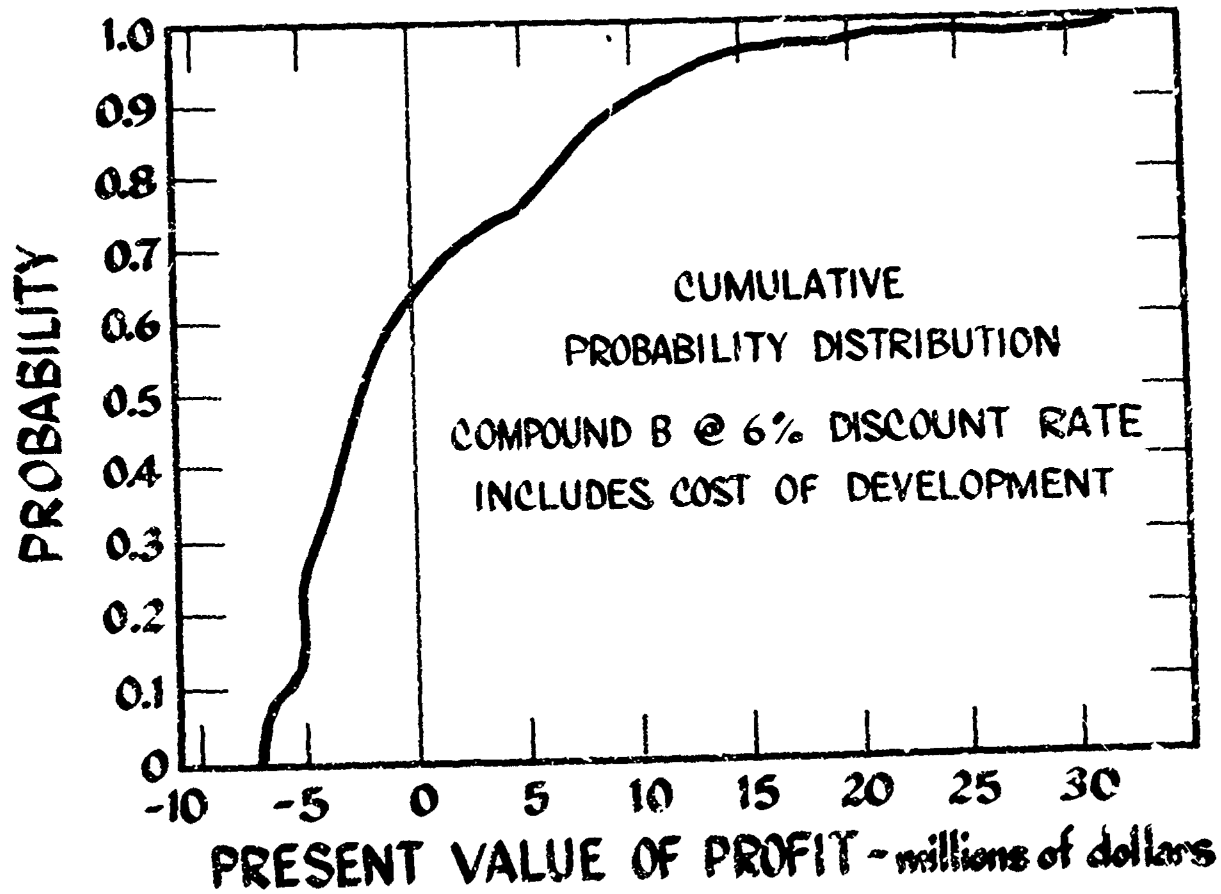


FIG 4 PROFIT LOTTERY FOR THE DEVELOPMENT OF COMPOUND B

to go ahead with the development of compound B in light of this new information. As a result, the client undertook a more extensive analysis of the international market for his product.

Space Program Planning

The space program planning application was conducted for the purpose of developing a methodology that would be useful in approaching technically complex decision problems; the intent was to carry out research on decision analysis itself. Although a very detailed analysis of the U.S. program for the unmanned exploration of Mars was conducted, no attempt was made to recommend specific decisions to the U.S. government. Instead, a large corporation that was quite familiar with the space effort played the role of the decision-maker during the analysis.

The problem was to determine the sequence of designs of rockets and payloads that should be used to pursue the goal of exploring Mars. It was considered desirable to place vehicles in orbit around Mars as well as to explore its atmosphere and to land vehicles on the surface of the planet to collect scientific data.

For purposes of obtaining sufficient information to encode properly the complex structure and information required to analyze this problem, a decision analyst resided with the client for a period of about one year. The client and the decision analyst worked as a team in building the models and submodels for the analysis.

The work was begun with a pilot phase, in which a simplified version of the decision problem was constructed. During this phase, four possible designs were postulated; each design represented increasing levels of sophistication. Figure 5 shows these designs and their potential

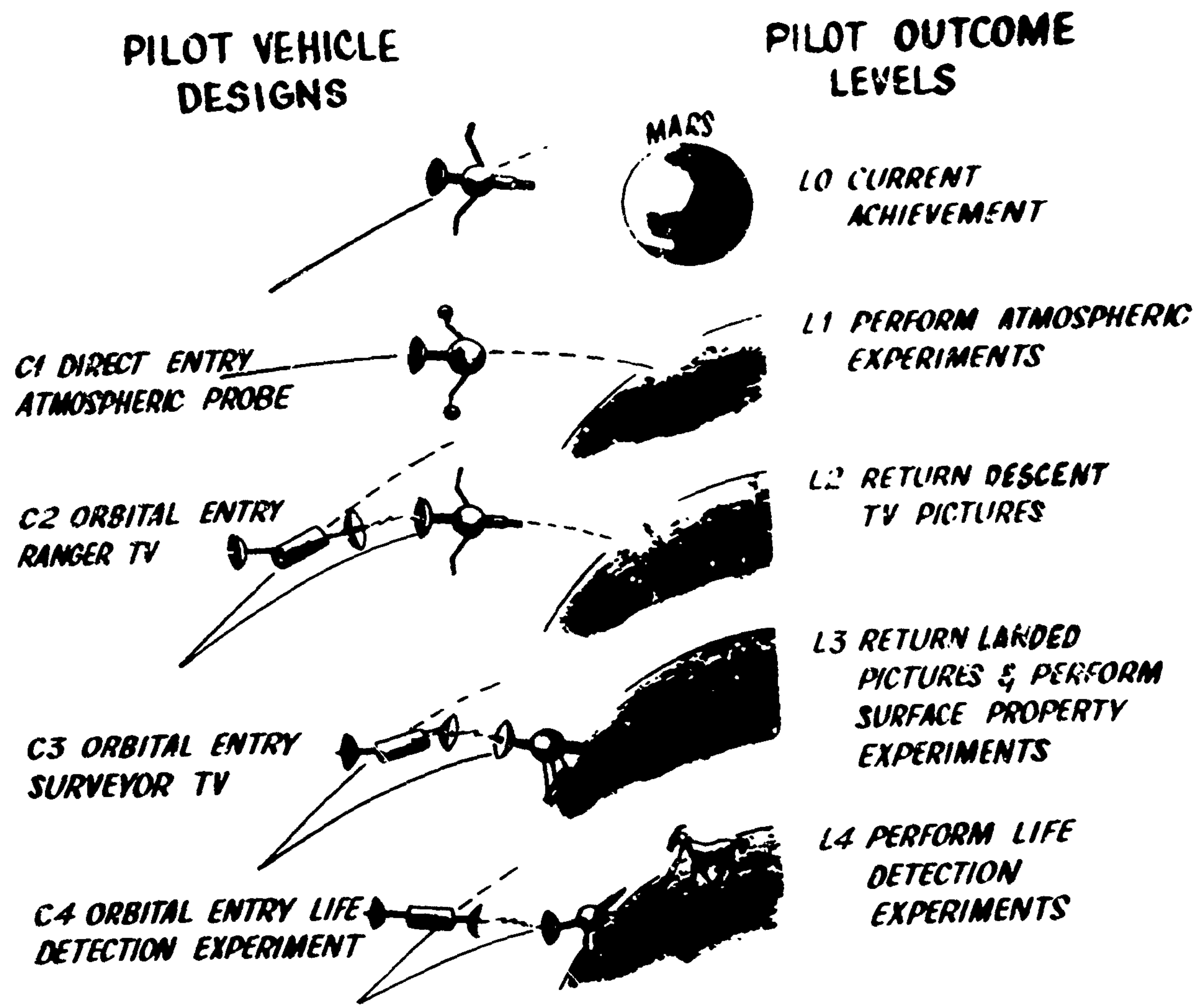


FIG 5 PILOT MODEL ASSUMPTIONS FOR THE EXPLORATION OF MARS

accomplishments. In the prototype analysis there were 12 possible vehicle designs plus the alternatives of skipping opportunities on cancelling the program at any decision point.

Because of the behavior of the orbits of the Earth and Mars, an opportunity to launch a vehicle toward Mars occurs about once every two years. Consequently, the decision problem was characterized by a sequential decision process, where each decision can be contingent upon the entire project history that precedes the decision point. Because of the lead time required in constructing a given vehicle, it was necessary to make each vehicle design decision before the outcome of the previous vehicle's flight was known. A decision tree was constructed to capture the structure of this sequential decision process.

In order to create a decision tree of manageable size, the concept of state variables was introduced. The state variables are a set of variables that are selected during the modeling process and whose value at any point in time summarizes all of the past history of the project relevant to future decision-making. Each node in the decision tree is characterized by a set of values for each of the state variables. The probabilities, cost, and values of subsequent branches are assigned conditionally on the basis of these values. Creativity is required in the selection of state variables. If a good approximation to the total available information is to be obtained, an appropriate set of state variables must be judiciously selected. A major objective in this process is to discover where essentially the same point can be reached via different paths through the program. When such a point is reached, two or more branches in the decision tree coalesce at a single node. The

node is assigned the common value of those state variables that are reached at this point along either path. This property, called coalescence, greatly reduces the size of the decision tree characterizing the problem. The sizes of the uncoalesced and coalesced decision trees for both the pilot and prototype decision trees are presented in Table 1.

The assignment of the probabilities, costs, and value parameters to the branches of the decision tree was a task that required the incorporation of information from additional submodels. For the pilot analysis, these models were kept quite simple.

In the prototype analysis, the most complex submodel was the probability model. Essentially, a probability tree was constructed from detailed diagrams that showed the functional steps in any flight to Mars. This tree had on the order of one hundred nodes, and the probabilities assigned to its branches were either obtained directly, from experimental judgment combined with experimental data, or indirectly from yet another sublevel of probability models. At each chance node in the decision tree, the detailed probability model produced the probability for each possible outcome.

Another unusual model was the value model, that is, the model that assigned a monetary value to each outcome in the space program. Since the client was reluctant to assign values directly in monetary terms, a cardinal scale of benefits was first employed. This scale was constructed so that the benefit of a perfect project would be one point. A total monetary value assignment to a perfect program then determined the monetary values to be used in the decision tree.

Table 1

SUMMARY OF DECISION TREE SIZE

	PILOT		FULL SCALE	
	Uncoalesced	Coalesced	Uncoalesced	Coalesced
Number of Nodes	3,619	56	476,012,807	3,153
Number of Branches	3,618	126	476,012,806	[illegible]2,784
Number of Paths	1,592	1,592	354,671,693	35[illegible]71,693
Number of Policies	3,005	somewhat less	over 10^{39}	somewhat less

The benefit scale was determined by constructing a value tree. The value tree is simply a convenient method of breaking the total benefit of the project into the incremental benefit of each individual outcome. Figure 6 shows a value tree for the pilot analysis. The value tree was constructed by dividing the benefit of the entire program (one point) into major categories, and then into subcategories identified in increasing detail until no further distinction is desirable. Each tip of this tree is divided into additional categories. Each additional category represents an elemental outcome that may be achieved during the project. For example, in the figure, the number 1.0 beside the node at the extreme left represents the total benefit of all the objectives of the program (achieving outcome L1, L2, L3, and L4 of Figure 5). The upper branch represents all direct scientific benefits of the program and was assigned 62% of the total value. The succeeding biological branch was assigned 60% of the scientific benefit, yielding 37% as the total benefit of the project to biological science. The further subdivision from this node represents the four increments in outcome level that are represented in Figure 5. Finally, the terminal node benefits were added for each level of outcome to give the totals shown in Figure 6. These totals, when multiplied by the total monetary value assigned to the program, determined the assignment of values to each outcome branch in the decision tree. A more detailed value tree was constructed for the prototype analysis.

In the pilot phase, calculations for the decision tree and the three submodels were made on a time-sharing computer system. The programming was carried out primarily by the decision analyst during the formation of the conceptual structure of the problem in the pilot phase.

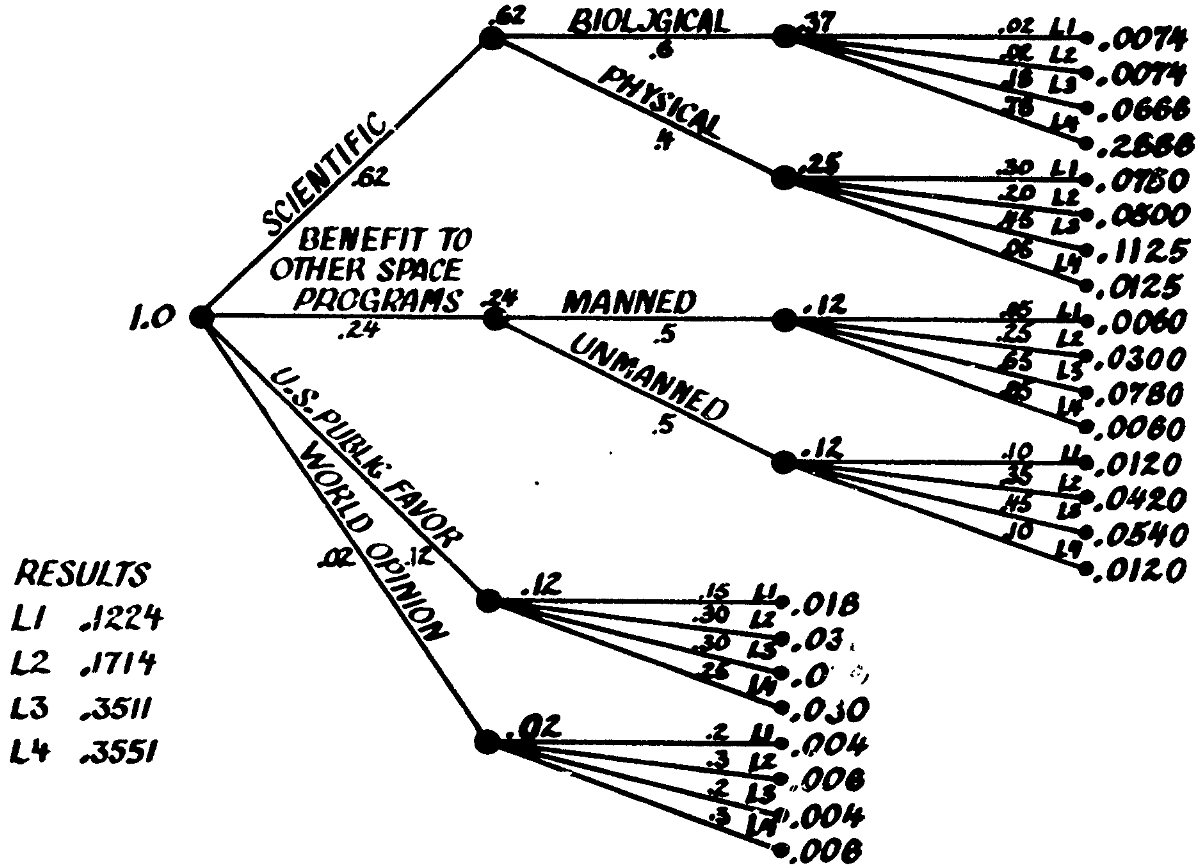


FIG 6 SPACE PROGRAM VALUE TREE

The pilot model provided a good means of communicating the concepts of the analysis and for making rapid sensitivity calculations. The pilot analysis could be carried out during meetings and presentations at which the results of changes in the parameters of the model could be determined almost instantaneously. In many cases, decision-makers would supply their values assignments for purposes of determining how the policy would be changed by them.

Because of the large size of the prototype model, the analysis programs were implemented in a system of programs called SPAN (Space Program ANalysis). The SPAN system is outlined in Figure 7.

The large size of the decision tree structure made it impractical to draw the complete tree by hand. Thus, the tree was generated by a computer program that utilized structural information describing characteristics of the decision tree to generate a symbolic description of the decision tree. This symbolic description was then compiled into a computer representation more suitable for computation. The generation and compilation were carried out in Phase 1.

In Phase 2 the cost, value, and probability model were executed, and from them, the numerical values of these parameters were generated and collated with the symbolic representations produced in Phase 1.

Phase 3 was a computer bookkeeping phase that operates on the decision model structure and the parameter tables for purposes of changing the information into a more efficient format for the analysis programs.

Phase 4 executed analysis programs that performed the roll-back of the decision tree, to determine optimum policies, and the

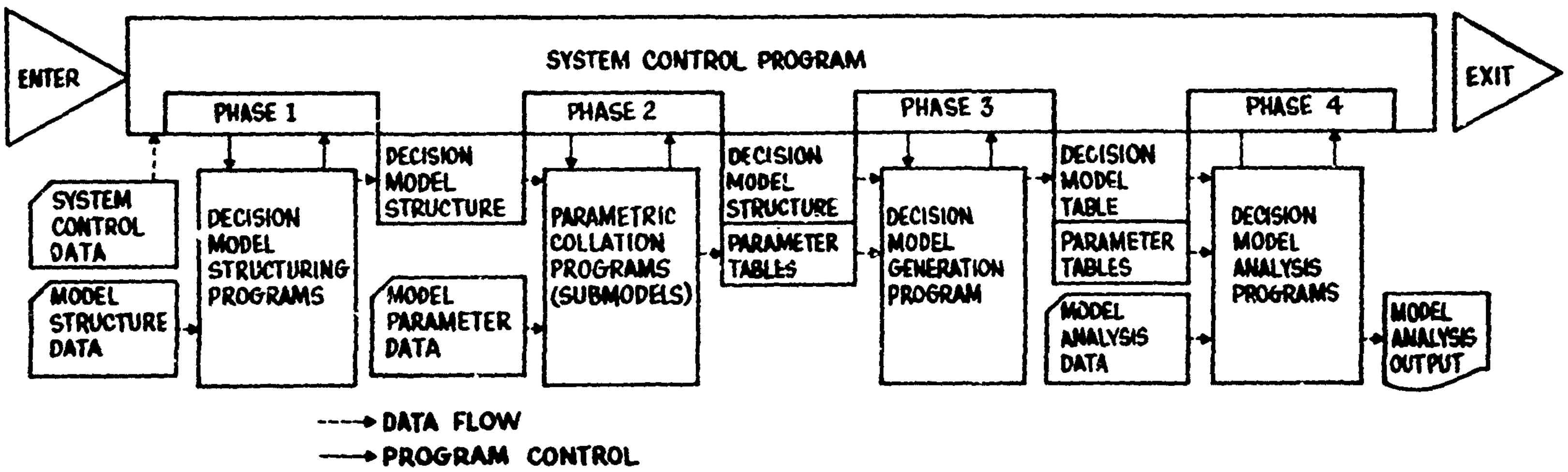


FIG 7 SPAN SYSTEM OPERATION

determination of the probabilities of the various events in the tree. It was capable of applying discount factors that represented time preferences and the exponential utility function that represented risk preference.

Electrical Power System Planning

The goal of this application was to create a basis for deciding when and whether to install a nuclear generating plant in Mexico. Because electrical generating plants have very long lifetimes, the desirability of any installation depends on the characteristics of the future system expansion. Consequently, each specific installation decision must be made within the framework of a policy for overall power system expansion.

In order to carry out this analysis, a project team, which included four representatives from Mexico and four decision analysts, was brought together for a period of about one year. The role of the Mexican representatives was to provide technological expertise, to collect necessary data, and to gather judgments regarding the preferences of the country of Mexico.

The conceptual framework for this problem is presented in Figure 8. At the left of the figure are the environmental inputs of the power system. These divide into four major categories--finance, energy, technology, and market. The financial model characterizes the terms at which capital is available from both domestic and world financial institutions and markets, as a function of the profitability, debt, and equity of the power utility. The energy model describes the price of all potential fuels--such as oil, natural gas, and uranium--as well as the availability of other energy sources--such as water power--over

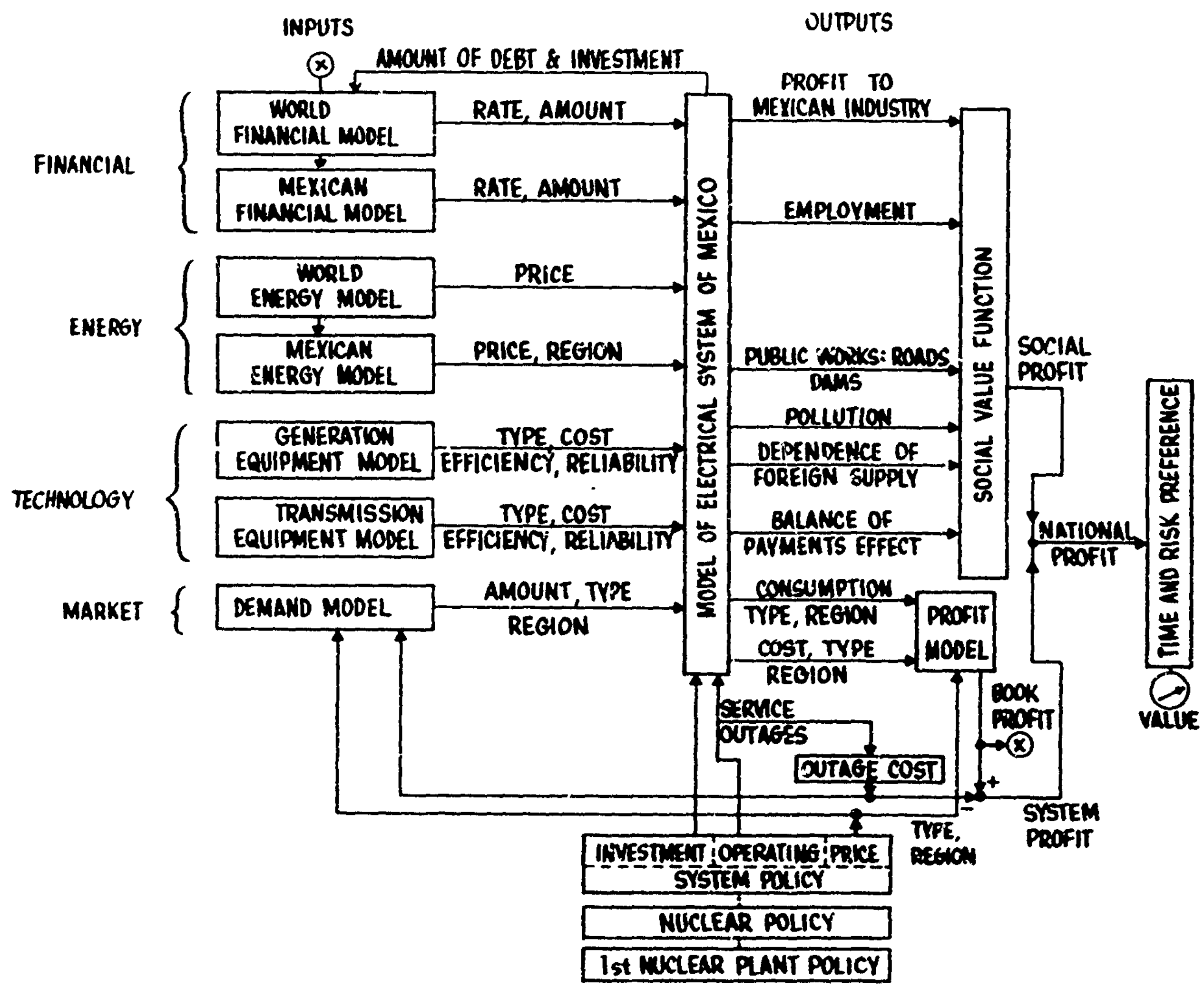


FIG 8 A DECISION ANALYSIS MODEL OF THE MEXICAN ELECTRICAL SYSTEM

the time period considered in the analysis. Similarly, the technology model characterizes the availability and prices of various types of generation and transmission equipment. Finally, the demand model describes the characteristics of electrical demand growth over time, ideally as a function of the price charged for electrical service.

At the bottom of the figure is the policy stating the conditions under which the first nuclear plant should be installed. The figure shows that this policy must be embedded in the general nuclear policy, which in turn is embedded in system's investment, operating, and pricing policy.

All of the environmental inputs and the policy alternatives feed into a model of the electrical system of Mexico. Application of the model determines the output variables over time. In the lower right corner of the figure, the outputs that indicate financial performance are shown. The amount of electrical consumption, the price of electricity, and the various costs are all combined to produce the usual book profit. Since reliability of service is one of the major considerations in electrical system expansion, the outage cost model is used to determine a monetary deduction from book profit, which yields system profit.

The social value function in the upper right-hand corner of the figure was included so that national goals that are outside the normal purview of the electrical system management could be considered. Its purpose is to assign a monetary value, called social profit, to social benefits of profit to Mexican industry, employment, public works, pollution, dependence on foreign supply, and effect on balance of payments. The sum of the social profit and the system profit is the national profit.

The uncertain time profile of national profit is converted into a single value, which might be called certain present national profit, by means of the time and risk preference model. The best decision policy is the one that maximizes the setting on this "value meter."

The development of this conceptual structure into a formal planning t[illegible] for system expansion proceeded through the pilot, prototype, and production stages described earlier. It must be pointed out that since an electrical system is so complex, different features of the planning model become important for different installation decisions. Thus, it is crucial that the analyst revalidate the model, through techniques such as sensitivity analysis, to ensure that it adequately captures the essence of each new installation decision.

The analysis was carried out through the development of a system of computer programs that simulate and evaluate the installation and operation of the electrical system over many years. The programs determine the cost of operation, including effects of maintenance, plant mix, system reliability, and possible energy deficits. Within this large simulation of the electrical system, the installation policy routines carried out less detailed simulations and evaluations of the system's future for the purpose of determining the time that each installation should be made and the type of installation it should be. The installation policy was refined so that the resulting installations would maximize the reading on the "value meter."

The pilot phase demonstrated the need for elaborate models that were capable of capturing the complexities of the electrical system problem. Thus, during the prototype phase, a modular system of computer

programs was constructed. This modular system facilitated the implementation of changes that would naturally occur in the transition to the production phase; and so that the appropriate module could be easily updated as the nature of the electrical system changes in the future. The computer model was constructed from a number of independent submodels that communicate through well-defined variables and tables.

One of the most significant submodels developed was the reliability submodel. In the ordinary expansion of an electrical system, each new plant is installed for the purpose of maintaining reliability in the face of demand growth. If plants did not randomly fail, an electrical system could operate with a much smaller capacity. Thus a computational procedure was developed to compute the system reliability from probabilistic demand information and the failure probabilities of each plant in the system. The effect of scheduled plant maintenance on reliability was included in the computation.

An interesting feature of power system expansion is that the system is self-healing. That is, if a "wrong" plant is installed at any time, or if the environment changes, the effects can be largely compensated for by the choice of new installations. Because an electrical system operates with a mix of plants--some best for steady base load and some best for rapid peaking--the new installations required by the usual rapid system growth can be selected so that the plant mix will be readjusted within a few years.

Gaps in the Theory

Perhaps the widest gaps between theory and practice are in the areas of values and preferences. Methods of solving even the seemingly simple

problem of characterizing time preference leave much to be desired. There is a great deal of controversy over the choice of a discount rate, and few guides exist for determining when a discount rate adequately represents time preference characteristics. Suggestions conflict about when the discount rate is used to represent actual time preference, when it is used to represent financing terms, and when it is sometimes even used to represent risk aversion.

Utility theory provides an elegant foundation for describing attitudes toward risk. However, seldom, if ever, are all the sources of uncertainty quantified. In addition, since each decision problem is part of a higher-level system, it is often not clear just what risk preference can be normatively deduced from higher-level considerations. In many applications, sensitivity to risk preference can be determined through the use of a family of utility functions, such as the exponential family.

Problems dominated by time or risk preference alone, usually can be adequately treated in spite of the above mentioned problems. However, when time and risk preference must be treated jointly, theoretical foundations are almost nonexistent. Techniques combining discount rates and the exponential family of utility functions were developed for use in the decision trees of the space program planning example.[5] A recent doctoral dissertation considers the joint time-risk preference from fundamental attitudes toward consumption.[6]

Some of the most perplexing problems arise, however, in the analysis of public decision problems. In the electrical system planning example, the space program planning example, and in applications to regulatory and natural resource decisions currently in progress, the

specification of the value function is a difficult task. The economic literature provides little guidance in the establishment of values for public decisions. In fact, many authors begin their developments with different implicit assumptions about the nature of the values. One example is the literature on marginal cost pricing.[7] I suspect that the resolution of these difficulties will come when the needs for explicit choices of public values are separated from their theoretical consequences.

Conclusion

The new discipline of decision analysis has been illustrated in practice with several examples. In my experience, decision analysis has proven to be a useful approach to complex decision problems. It provides not only the principles necessary for analysis, but also a means of bringing the important issues of the problem into focus, so that new alternatives can be created, information gathering possibilities can be evaluated, and the analysis effort itself can be efficiently channeled. Applications have shown the need for new theory and methodology for treatment of values and preferences, especially in public decision problems.

References

[1] R. A. HOWARD (1966) Decision Analysis: Applied Decision Theory. Proceedings of the Fourth International Conference on Operational Research, Session 2, International Federation of Operational Research Societies, Boston.

[2] R. A. HOWARD (1968) The Foundations of Decision Analysis. IEEE Transactions on Systems Science and Cybernetics. Vol. SSC-4, 3.

[3] J. E. MATHESON (1968) The Economic Value of Computation and Analysis. IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, 3.

[4] J. E. MATHESON and W. J. ROTHS (1967) Decision Analysis of Space Projects. Proceedings National Symposium Saturn/Apollo and Beyond American Astronautical Society, June 11, 1967.

[5] J. E. MATHESON and R. A. HOWARD (1967) Decision Analysis of Multifurcate Trees. Presented at the Institute of Management Science XIV International Meeting, Mexico City, August 24, 1967.

[6] A. POLLARD (1969) Ph.D. dissertation (in preparation), Department of Engineering Economic Systems, Stanford University.

[7] N. RUGGLES (1949-50) Recent Developments in the Theory of Marginal Cost Pricing. The Review of Economic Studies, Vol. 17.

A CASE HISTORY OF RESEARCH THAT FAILED

by

RALPH O. SWALM
Syracuse University

Businessmen have long characterized themselves as takers of risks; at long last some of them are now beginning to use analytical methods that explicitly recognize this fact. And more and more schools that produce tomorrow's business leaders speak, in their new jargon, of building risk rather than certainty models to aid in decision making processes.

Increasingly, it is apparent that there are better ways to recognize the risks inherent in business than to shorten the payout period, or to increase the minimum attractive rate of return required to justify an investment. In our schools, and in practice, we are coming to seek specific probability statements concerning possible future events, and to take these into consideration when making decisions. We still have a lot to learn regarding ways and means of making better probability assessments, and we need to look much more closely at the criteria we wish to use once these assessments are made.

One widely used criterion is to assume that a businessman should choose that strategy that optimizes the expected value of the outcome. (Expected value is a statistician's word for a value you almost certainly don't expect-in simple language, the expected value of a strategy is its average value if that strategy is repeated in the same situation an infinitely large number of times.)

It is perhaps intuitively obvious that if a situation is encountered an infinite number of times, one can do no better than to choose that strategy that gives the greatest average gain per time-provided one does not go broke in the meantime. But the only way one can be assured of not going broke if there is a finite probability of loss each time is by having infinite resources! But most interesting decisions are made only once-and few are so fortunate as to have unlimited resources. What then?

Well, if the amount risked is small relative to one's resources, most people would be willing to use the maximum expected value criterion even on one-time opportunities. For example-would you not be willing to toss a coin if you would receive five guilders if a head showed, but would have to pay one guilder if a tail came up? The "expected value" of this game is 1/2 x 5 + 1/2 x -1, or two guilders. Even if this game cost one guilder to play, its net expected value would be one guilder, and many would use this information in deciding to play.

But before you conclude that maximizing expected value is a pretty good way to make decisions, consider this proposition. You have two choices. One is to accept a million tax free guilders and quit. The other is to toss a coin - if heads show, you get five million tax free guilders, but if a tail shows you get nothing. Obviously, the expected value of your second option is 1/2 x 5 million + 1/2 x 0, or 2 1/2 million; the first option pays only 1 million. But wouldn't you take the sure one million anyway? Most others would.

What this simple example shows is that in at least one case, most people will not optimize on expected value. And it is not difficult to show that this is true in a large class of cases, particularly in those in which a probability of a significant loss is present and recognized.

Corporate, as well as of individual, decision makers can also be shown to be unwilling to select the course of action that optimizes the expected value of the outcome in many cases. How, then, can we hope either to predict or to prescribe which of several risky options a businessman would, or should, select?

The more interesting question is how businessmen should select among risky projects, I would argue that, as a minimum, various decision makers within a company should: (1) In an as explicit way as is possible, discuss desireable risk attitudes, so they can (2) Define a corporate attitude toward risk, and (3) Agree, individually and collectively, to take that corporate attitude into consideration in reaching decisions.

This, of course, is not to say that corporate attitudes toward risk must remain stable over long periods of time nor that any decisions should be made on a purely mathematical basis.

I have spoken of the way that corporate decision makers should participate in risk decisions. How do they behave?

The hard evidence on which to base an answer to this question is all too limited, but it indicates quite clearly that here, as in so many other areas, how they do behave is a far cry from how they should behave.

May I share with you some of the evidence upon which this statement is based? A short time ago, I had the privilege of studying the risk attitudes of every decision maker, from foreman to chairman of the Board of Directors, of a company employing about six hundred people. To do this, I asked each of them, in individual sessions how he would recommend that a series of decisions involving risks be made.

The twenty-eight decision makers I queried gave startling diverse replies. For example, one foreman said that he would be indifferent between recommending an automation project that promised a 50-50 chance of earning a net of either $30,000 or nothing or a sure thing investment that would net $15,000; another, when asked the same question, said "I'd sure like to recommend the automation project over a hand method that

would give a sure $5,000, but as a company man I can't." He finally concluded that he would be indifferent between recommending automation equipment offering a 50-50 chance of netting either $30,000 or nothing and a hand method promising a certain $1,200.

And the upper levels, the President indicated indifference between a research project that would offer a 50-50 chance of netting either $200,000 or nothing as compared to a development project promising a sure $30,000. The board chairman's indifference point for the same gamble was a sure $100,000.

Directly contrasting the risk attitudes of the four men mentioned one was indifferent between a 50-50 chance of the company's making either $30,000 or nothing and a sure $1,200, one a sure $8,000 and two a sure $15,000. While these four replies include both ends of the spectrum, others pretty well covered all intermediate possibilities.

A similar range of replies was found in questions in which a possibility of loss was contemplated. For example, one question was "You feel that you have a 50-50 chance of getting a certain contract. If you do, your company will net $200,000. If you do not get it, you will lose X dollars. At what value of X would you become indifferent to either recommending trying or not trying for the contract?" Respondents placed X as low as $1,200 and as high as $100,000.

Six men were asked specifically how great the potential gain would have to be for them to recommend a 50-50 gamble that could lose $20,000 for the company. Their replies were 50, 70, 80, 100, 150 and 200 thousand dollars.

As the research progressed, it became more and more clear that guidance was both wanted and needed in spelling out corporate risk attitudes. For example, in reply to one question, the sales manager told me, "That is easy. I know the corporate attitude on that sort of thing." He then went on to give me quite a different answer from that which the president had given me just a few hours earlier! One man found the questions almost impossible to answer, "because this company would never knowingly invest money if any risk were involved." Another claimed difficulty because, "I'm not convinced that many decisions are really gambles. If you just put in enough effort you can get the data so that no decision is really necessary. The facts will speak for themselves." Another, when asked, "Then you view company as conservative?", replied, "I do, extremely conservative. And I try to fit the pattern." Another said, "My God, if we don't take risks once in a while, we are dead!" And still another offered the comment, "Yeah, I know they'd like us to be conservative, but I don't know just how conservative."

Although they certainly didn't couch it in these terms, one of the factors frequently brought up by the respondents was the lack of information in single points as opposed to probabilistic predictions. A typical incident in which this occured was the following:

The tool designer showed me an order he had been given to make a mold produce 10,000 pieces per week of a certain piece. Turning to me, he said, "Do you know what I'm going to do? Instead of making the five cavity mold that 10,000 pieces a week would require, I'll make a mold with spaces for five cavities but build only three of them. I have found through bitter experience how optimistic those salesmen are!"

We went on to discuss what he would do if given orders for molds for two similar items. The first was a par for which there were firm orders for 10,000 per week. The second was a new design for which the best estimate was 10,000 per week but which might range anywhere from twice that figure to nothing. He indicated that, in the first case, he would build a five cavity mold and be done with it; in the second, he would complete perhaps three cavities on an eight cavity mold. "Gosh", he said, "It would be easy if they just gave figures like that. But now, I just have to guess what they mean when they say 10,000 pieces a week will be required."

Now, I ask you-who is better able to establish the probability distribution of future sales-the sales department or the tool designer? And when this distribution is established, should the final decision on how to build the mold rest on the tool designer's risk attitudes or those of the company management? I know the designer's answer-he wanted help.

The original purpose of my study was not to find out how the various decision makers did behave-at that time I had ample evidence to predict with confidence that a wide disparity of risk attitudes would be found-but to show them this disparity and to suggest that choosing an appropriate attitude toward risk was a high corporate policy matter that should not be decided, as it were, on default. I wish I could offer you a happy ending to this story, describing how management reasoned together and came to a better agreement as to a consistent, rational risk attitude. But I can't. After some discussion, the foremen agreed that such a policy would be most helpful to them and they agreed that the optimal policy would be considerably less risk-averse than their present attitudes were. They justified the discrepancy between what they thought their attitude should be and what it was by noting that they considered their company to be overly conservative, and tailored their recommendations accordingly. The president stated that he was aware these feelings existed, but felt that it was truly not a conservative company. He pointed out that few good proposals had been turned down at top level as evidence of this fact.

But when your author pointed out the vast differences in risk attitudes among his staff and pointed out the desire of the lower management levels for guidance in formulating risk attitudes, then tried to emphasize the need for a reasonable attempt to at least discuss the formulation of a corporate risk policy among the members of his "cabinet", he met with no success. The President persisted in his belief that perhaps diversity was good, and quite properly insisted on his right to terminate the study then and there without any further explanation. I fear this result speaks poorly of your author's ability as a salesman, but I must, in truth, report failure at this critical point.

Why, then, do I write this sad tale? Because I feel that what I tried to do needs doing, and perhaps the reader can succeed where the author failed. Because the need for a corporate policy was articulated by so many in this corporation that I am utterly convinced that it is real in this and other corporations, and that this need poses us all a real challenge to meet it.

A THEORY OF IDEAL LINEAR WEIGHTS FOR
HETEROGENEOUS COMBAT FORCES*

David R. Howes and Robert M. Thrall

Introduction. In conducting military research, analysts frequently make use of indices of force effectiveness which attempt to describe the value of the force to its side in some hypothetical military conflict. Firepower potentials are an example of such indices. Current study of the problem suggests that indices can perhaps more effectively be based on effectiveness matrices, such as might emerge from a detailed combat simulation or from other sources.

When such tables are given it is possible to construct from them a system of weapon weights each of which is a weighted average of the effects of a given weapon against each of the enemy's weapons. This paper will describe the construction of such weights.

1. Effectiveness matrices. In military combat, the only tangible, quantifiable value of a weapon system, as opposed to some other type of system, is expressed in terms of the damage which it produces. A weapon may act through its capacity to deny an enemy certain tactical options, however its final quality is lethality. Weapon effectiveness may be considered a function of casualty-production which lies in depriving the enemy of the value of weapons lost. Therefore, it is appropriate to consider numbers which measure the killing power of each weapon against each opposing weapon. An effectiveness matrix may be regarded as a table whose entries are these killing powers or relative effectivenesses.

More precisely, consider a combat situation between two opponents, Blue and Red. We suppose that Blue has m classes of weapons and consider the Blue force vector

*A working paper RMT-200-W13-28R, Robert M. Thrall and Associates, March, 1972.

Preceding page blank

(1.1) $$U_B = \begin{bmatrix} u_{1B} \\ \cdot \\ \cdot \\ \cdot \\ u_{mB} \end{bmatrix}$$

where u_{1B} is the number of Blue weapons of class $1,\ldots,u_{mB}$ is the number of Blue weapons of Blue class m. Similarly, suppose that R has n classes of weapons and that

(1.2) $$U_R = \begin{bmatrix} u_{1R} \\ \cdot \\ \cdot \\ \cdot \\ u_{nR} \end{bmatrix}$$

is the Red force vector.

We wish to find Blue and Red <u>weight vectors</u>

(1.3) $$W_B = \begin{bmatrix} w_{1B} \\ \cdot \\ \cdot \\ \cdot \\ w_{mB} \end{bmatrix} , \quad W_R = \begin{bmatrix} w_{1R} \\ \cdot \\ \cdot \\ \cdot \\ w_{nR} \end{bmatrix}$$

such that the linear combinations

(1.4) $$S(B) = w_{1B}u_{1B} + \ldots + w_{mB}u_{mB} = W_B^T U_B$$

and

(1.5) $$S(R) = w_{1R}u_{1R} + \ldots + w_{nR}u_{nR} = W_R^T U_R$$

are <u>good</u> measures of the respective overall strengths of Blue and Red. Then the fraction

$$T = S(B)/S(R) \tag{1.6}$$

can be used as an index, called the THOR Index, [see Reference 4] of the relative strengths.

A Blue-vs-Red effectiveness matrix M_{BR} is a matrix (table) having m rows and n columns where the element $m_{BR}(i,j)$ measures the effectiveness (killing power) of a single weapon of Blue class i against Red weapon class j. Similarly a Red-vs-Blue effectiveness matrix

$$M_{RB} = [m_{RB}(j,i)] \tag{1.7}$$

has n rows and m columns and, inversely, $m_{RB}(j,i)$ measures the effectiveness of a single Red weapon of class j against Blue weapon class i. The numbers $m_{BR}(i,j)$ and $m_{RB}(j,i)$ may be positive or zero but, by definition, cannot be negative.

For example, suppose that $m = n = 2$, that both Red and Blue weapon class one is an infantry company and that both Red and Blue weapon class two is an artillery battery. Then the effectiveness matrices

$$M^1_{BR} = \begin{bmatrix} .5 & 0 \\ .7 & .2 \end{bmatrix}, \quad M^1_{RB} = \begin{bmatrix} .6 & 0 \\ .6 & .1 \end{bmatrix} \tag{1.8}$$

would describe a situation in which (1) in infantry combat Red was more effective than Blue (.6 vs .5), (2) neither infantry could harm the enemy artillery, and (3) the Blue artillery is superior to the Red artillery, and (4) each artillery battery has a positive effectiveness against its counterpart.

The effectiveness matrices

$$M^2_{BR} = \begin{bmatrix} .5 & .1 \\ .7 & .2 \end{bmatrix}, \quad M^2_{RB} = \begin{bmatrix} .6 & .2 \\ .6 & .1 \end{bmatrix} \tag{1.9}$$

would describe a change which gave each infantry capability against the opposing artillery.

The matrices

$$(1.10) \qquad M_{BR}^{3} = \begin{bmatrix} .5 & 0 \\ .7 & .8 \end{bmatrix}, \quad M_{RB}^{3} = \begin{bmatrix} .6 & 0 \\ .6 & .5 \end{bmatrix}$$

would describe a different type of change in which the artillery attritions are substantially increased.

If we assume that the artillery units are either concealed or out of each other's range then we could have effectiveness matrices

$$(1.11) \qquad M_{BR}^{4} = \begin{bmatrix} .5 & 0 \\ .7 & 0 \end{bmatrix}, \quad M_{RB}^{4} = \begin{bmatrix} .6 & 0 \\ .6 & 0 \end{bmatrix}.$$

2. Ideal linear weights. We turn next to consideration of suitable weight vectors W_B and W_R. These should be derived in some reasonable way from the corresponding effectiveness matrices M_{BR} and M_{RB}.

For example, one could simply let W_B be the average of the columns of M_{BR}. Using M_{BR}^1 and M_{RB}^1 this would give

$$(2.1) \qquad W_B^1 = \tfrac{1}{2}\begin{bmatrix} .5 + 0 \\ .7 + .2 \end{bmatrix} = \begin{bmatrix} .25 \\ .45 \end{bmatrix}, \text{ and } W_R^1 = \begin{bmatrix} .3 \\ .35 \end{bmatrix};$$

similarly from M_{BR}^2 and M_{RB}^2 we would obtain

$$(2.2) \qquad W_B^2 = \begin{bmatrix} .3 \\ .45 \end{bmatrix}, \quad W_R^2 = \begin{bmatrix} .4 \\ .35 \end{bmatrix}.$$

This naive approach has the advantage of simplicity, but lacks credibility since it places equal emphasis on effectiveness against enemy infantry and

artillery whereas one of these might be considered much more dangerous than the other.

The naive approach places equal weight on each column. A more general procedure is to select as weights non-negative numbers which add to one. Thus in example 2, if we consider enemy artillery to be twice as important a target as enemy infantry we would choose weights 1/3, 2/3 and get

$$W_B^2 = \begin{bmatrix} \frac{1}{3}(.5) + \frac{2}{3}(.1) \\ \frac{1}{3}(.7) + \frac{2}{3}(.2) \end{bmatrix} = \frac{1}{3}\begin{bmatrix} .7 \\ .11 \end{bmatrix} .$$

A vector with non-negative elements that sum to one is called a probability vector. Then the more general procedure would consist of selecting two probability vectors

$$(2.3) \qquad Z_B = \begin{bmatrix} z_{B\,1} \\ \cdot \\ \cdot \\ \cdot \\ z_{B\,m} \end{bmatrix} , \quad Z_R = \begin{bmatrix} z_{R\,1} \\ \cdot \\ \cdot \\ \cdot \\ z_{R\,n} \end{bmatrix}$$

and then defining the linear weights by

$$(2.4) \qquad W_B = M_{BR} Z_R \quad , \quad W_R = M_{RB} Z_B$$

The next step is selection of Z_B and Z_R. In the naive approach we took

$$(2.5) \qquad Z_B = \frac{1}{m} E_m = \frac{1}{m}\begin{bmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} , \quad Z_R = \frac{1}{n} E_n = \frac{1}{n}\begin{bmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} .$$

Here (and later) we use the symbol E_p to represent the column vector

consisting of p ones, e.g.

$$E_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} .$$

A second, somewhat more reasonable selection is

(2.6) $$Z_B = M_{BR} E_n / \gamma_B \quad , \quad Z_R = M_{RB} E_m / \gamma_R$$

where

(2.7) $$\gamma_B = E_m^T M_{BR} E_n = \sum_{i,j} m_{BR}(i,j) \; , \; \gamma_R = E_n^T M_{RB} E_m = \sum_{i,j} m_{RB}(i,j)$$

then

(2.8) $$W_B = M_{BR} M_{RB} E_m / \gamma_R \quad , \quad W_R = M_{RB} M_{BR} E_n / \gamma_B \; .$$

In Example 2 this gives

(2.9) $$Z_B^2 = \begin{bmatrix} .6 \\ .9 \end{bmatrix} / 1.5 \quad , \quad Z_R^2 = \begin{bmatrix} .8 \\ .7 \end{bmatrix} / 1.5$$

and

(2.10) $$W_B^2 = M_{BR}^2 Z_R^2 = \begin{bmatrix} .47 \\ .70 \end{bmatrix} / 1.5 \quad , \quad W_R^2 = \begin{bmatrix} .54 \\ .45 \end{bmatrix} / 1.5 \; .$$

These procedures are only two among many possibilities for choosing the probability vectors Z_B and Z_R . The one which we next introduce and recommend for serious consideration yields weights W_B and W_R which we call <u>ideal linear weights</u>.

To motivate them we consider the following argument. Suppose that W_R has been determined; this means that relative values for the Red weapon systems are known. Then it seems reasonable to select as Z_R the unique probability vector proportional to W_R . Similar reasoning would apply in selection of Z_B if W_B is given. This line of argument would lead to

$$(2.11) \qquad Z_B = W_B/\alpha_B \quad , \quad Z_R = W_R/\alpha_R$$

where $\qquad \alpha_B = E_m^T W_B \quad , \quad \alpha_R = E_n^T W_R$

then, we get

$$(2.12) \qquad W_B = M_{BR} Z_R = M_{BR} W_R/\alpha_R \quad , \quad W_R = M_{RB} Z_B = M_{RB} W_B/\alpha_B$$

and substituting each of these equations in the other we get

$$(2.13) \qquad W_B = M_{BR} M_{RB} W_B/\alpha_B \alpha_R \quad , \quad W_R = M_{RB} M_{BR} W_R/\alpha_R \alpha_B \quad .$$

Now, let

$$(2.14) \qquad P_B = M_{BR} M_{RB} \quad , \quad P_R = M_{RB} M_{BR} \quad , \quad \lambda = \alpha_B \alpha_R$$

and we have the equations

Reproduced from best available copy.

$$(2.15) \qquad P_B W_B = \lambda W_B \quad ; \quad P_R W_R = \lambda W_R \quad .$$

The ideal weights must satisfy these equations and also be non-negative vectors (and also non-zero).

Fortunately, these equations are well known in linear algebra. First, they require that λ be an eigenvalue of each of the square matrices P_B (m by m) and P_R (n by n) and that W_B, W_R be eigenvectors. Since the effectiveness matrices M_{BR}, M_{RB} have non-negative elements the same is true of their products P_B, P_R.

The classical Perron-Frobenius theory of eigenvalues and eigenvectors of non-negative matrices applies to our situation and guarantees solutions to (2.14) with W_B, W_R non-negative and λ positive. Moreover, it follows from the general theory of matrices that P_B and P_R have the same non-zero eigenvalues. The pertinent facts from the classical Perron-Frobenius theory

can be found (with proofs) in Chapter XIII of Gantmacher, Vol II [see Reference 2]. This chapter also has a comprehensive bibliography. The original papers by Perron and Frobenius appear, respectively, as References 3 and 1 below.

3. Examples of ideal weights. We return to our four examples to illustrate the theory.

Example 1.

$$P_B^1 = \begin{bmatrix} .30 & 0 \\ .54 & .02 \end{bmatrix} \quad , \quad P_R^1 = \begin{bmatrix} .30 & 0 \\ .37 & .02 \end{bmatrix} . \tag{3.1}$$

The eigenvalues for both P_B^1 and P_R^1 are $\lambda_1^1 = .30$, $\lambda_2^1 = .02$. Then

$$Z_B^1 = \begin{bmatrix} .34 \\ .66 \end{bmatrix} \quad , \quad Z_R^1 = \begin{bmatrix} .43 \\ .57 \end{bmatrix} \tag{3.2}$$

are the unique probability eigenvectors corresponding to λ_1^1 . The corresponding weights are

$$W_B^1 = M_{BR}^1 Z_R^1 = \begin{bmatrix} .215 \\ .415 \end{bmatrix} \quad , \quad W_R^1 = \begin{bmatrix} .204 \\ .270 \end{bmatrix} \tag{3.3}$$

$$\alpha_B^1 = .63 \ , \quad \alpha_R^1 = .474 \quad , \quad \alpha_B^1 \alpha_R^1 = \lambda_1^1 = .3 \quad .$$

The second eigenvalue λ_2^1 gives

$$Z_B^{1*} = Z_R^{1*} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} , \quad W_B^{1*} = \begin{bmatrix} 0 \\ .2 \end{bmatrix} \quad , \quad W_R^{1*} = \begin{bmatrix} 0 \\ .1 \end{bmatrix} . \tag{3.4}$$

We will see later that this second eigenvalue yields less meaningful weights than the first.

Example 2.

$$(3.5) \qquad P_B^2 = \begin{bmatrix} .36 & .11 \\ .54 & .16 \end{bmatrix} , \quad P_R^2 = \begin{bmatrix} .44 & .10 \\ .37 & .08 \end{bmatrix} .$$

The characteristic equation for both matrices is

$$(3.6) \qquad \lambda^2 - .52\lambda - .0018 = 0$$

and has as its roots the eigenvalues

$$(3.7) \qquad \lambda_1^2 = .5235 , \quad \lambda_2^2 = -.0035 .$$

From λ_1^2 we get the unique probability eigenvectors

$$(3.8) \qquad Z_B^2 = \begin{bmatrix} .40 \\ .60 \end{bmatrix} , \quad Z_R^2 = \begin{bmatrix} .545 \\ .455 \end{bmatrix}$$

for P_B^2 amd P_R^2 respectively .

$$(3.9) \qquad W_B^2 = \begin{bmatrix} .32 \\ .48 \end{bmatrix} , \quad W_R^2 = \begin{bmatrix} .36 \\ .30 \end{bmatrix}$$

$$\alpha_B^2 = .8 , \quad \alpha_R^2 = .66 , \quad \alpha_B^2 \alpha_R^2 = .528 \sim \lambda_1^2 .$$

Example 3.

$$(3.10) \qquad P_B^3 = \begin{bmatrix} .3 & 0 \\ .9 & .4 \end{bmatrix} , \quad P_R^3 = \begin{bmatrix} .30 & 0 \\ .65 & .40 \end{bmatrix} , \quad \lambda_1^3 = .3 , \quad \lambda_2^3 = .4 .$$

This example differs from Example 1 since this time the second eigenvalue is larger than the first. The only probability eigenvectors come from λ_2^3 and are

(3.11) $$Z_B^3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} , \quad Z_R^3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} .$$

These give

(3.12) $$W_B^3 = \begin{bmatrix} 0 \\ .8 \end{bmatrix} , \quad W_R^3 = \begin{bmatrix} 0 \\ .5 \end{bmatrix} , \quad \alpha_B^3 = .8, \quad \alpha_R^3 = .5, \quad \alpha_B^3 \alpha_R^3 = .4 = \lambda_2^3 .$$

Example 4.

(3.13) $$P_B^4 = \begin{bmatrix} .30 & 0 \\ .42 & 0 \end{bmatrix} , \quad P_R^4 = \begin{bmatrix} .30 & 0 \\ .30 & 0 \end{bmatrix} . \quad \lambda_1^4 = .3 , \quad \lambda_2^4 = 0 .$$

This example resembles Example 1 in that the first eigenvalue is larger than the second.

From the first eigenvalue we get

(3.14) $$Z_B^4 = \begin{bmatrix} .42 \\ .58 \end{bmatrix} , \quad Z_R^4 = \begin{bmatrix} .5 \\ .5 \end{bmatrix}$$

$$W_B^4 = \begin{bmatrix} .25 \\ .35 \end{bmatrix} , \quad W_R^4 = \begin{bmatrix} .25 \\ .25 \end{bmatrix} , \quad \alpha_B^4 = .6, \quad \alpha_R^4 = .5, \quad \alpha_B^4 \alpha_R^4 = .3 = \lambda_1^4 .$$

The second eigenvalue gives

(3.15) $$Z_B^{4*} = Z_R^{4*} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$W_B^4 = W_R^4 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} , \quad \alpha_B^4 = \alpha_R^4 = 0 ,$$

and thus does not provide useful weighting vectors.

Example 2 illustrates a general class of situations where each Blue weapon system is (at least minimally) effective against each Red one and vice versa. If a square matrix P has positive (not merely non-negative)

elements then it has a unique probability eigenvector Z and the corresponding eigenvalue λ_1 (called the Perron eigenvalue) is not only positive but has the largest absolute value of all the eigenvalues of P. It is then easy to calculate Z and λ_1 by the following sequential process. Let $V_0 = E_m$ (where P is m by m), let $\alpha(V_0) = E_m^T V_0 = m$, let $Z_0 = V_0/\alpha(V_0)$, and proceeding inductively let $V_{i+1} = PZ_i$, let $Z_{i+1} = V_{i+1}/\alpha(V_{i+1})$, $i = 1,2,\ldots$. Then

$$(3.16) \qquad Z = \lim_{i\to\infty} Z_i \quad , \quad \lambda_1 = \lim_{i\to\infty} \alpha(V_{i+1}) \; .$$

These results still hold even if P has some, but not too many, zero elements.

Indeed, when P_B and P_R are positive, we can use a limiting process to define the ideal weights W_B, W_R . .

We can begin with W_R^0 any positive vector (e.g., $W_R^0 = E_n$) then in turn set $Z_R^0 = W_R^0/\alpha(W_R^0)$, $W_B^0 = M_{BR} Z_R^0$, $Z_B^0 = W_B^0/\alpha(W_B^0)$, and proceeding inductively

$$(3.17) \qquad W_R^i = M_{RB} Z_B^{i-1} \quad , \quad Z_R^i = W_R^i/\alpha(W_R^i)$$

$$W_B^i = M_{BR} Z_R^i \quad , \quad Z_B^i = W_B^i/\alpha(W_B^i) \quad , \quad i = 1,2,\ldots \; .$$

Then the six sequences

$$(3.18) \qquad W_R^i , \quad Z_R^i , \quad W_B^i , \quad Z_B^i , \quad \alpha(W_R^i), \quad \alpha(W_B^i)$$

converge respectively to

$$(3.19) \qquad W_R , \quad Z_R , \quad W_B , \quad Z_B , \quad \alpha_R , \quad \alpha_B$$

where Z_B, Z_R are the unique Perron probability eigenvectors of P_B, P_R, respectively; W_R, W_B are the ideal weights for R, B respectively;

$$W_R = \alpha_R Z_R , \quad W_B = \alpha_B Z_B \quad \text{and} \quad \lambda_1 = \alpha_R \alpha_B$$

is the Perron eigenvalue for both P_B and P_R .

This approach provides a computationly convenient algorithm for calculating the ideal weights. When m and n exceed two this approach is clearly preferable to calculating and solving the characteristic equation for P_B or P_R . There are other more refined computational algorithms which are in general more efficient than this one. However, a computer program written for this iterative process gave quite satisfactory numerical results for moderate values of m and n. An example involving 40 weapon classes converged in 9 iterations to an accuracy of .0001.

4. Interpretation of reducibility.

Examples 1, 3, 4 illustrate some of the possible effects of zeros in P_B, P_R . All of the P's in these examples are what is called reducible . A non-negative square matrix P is said to be reducible if it has the form

$$P = \begin{bmatrix} P_1 & 0 \\ P_{21} & P_2 \end{bmatrix}$$

where P_1 and P_2 are square, or more generally, if this form can be obtained by reordering of the rows followed by the same reordering of the columns.

In our combat context, we encounter reducible matrices when as in Examples 1, 3, 4 there are two classes of weapons on each side and the first class of Blue is totally ineffective against the second class of Red and vice versa.

Let us assume that both P_B and P_R are reducible with P_{B1}, P_{B2}, P_{R1}, P_{R2} all positive, that P_{B1}, P_{R1} have the Perron eigenvalue λ_1 , and that P_{B2}, P_{R2} have the Perron eigenvalue λ_2 . [These assumptions all hold for Examples 1 and 3.] Then, if we apply our computational algorithm beginning with $W_R^0 = E_n$, the limiting eigenvectors obtained will correspond to the larger eigenvalue.

Thus, in Example 1 we would get W_B^1, W_R^1 and not W_B^{1*}, W_R^{1*}. In Example 3 we would, of course, get W_B^3, W_R^3 and in this case there is no possibility of positive ideal weights.

Moreover, in Example 1 the only way to get the starred vectors would be to start with W_R^0 of the form $\begin{bmatrix} 0 \\ a \end{bmatrix}$, i.e., almost all starting vectors W_R^0 will yield W_B^1, W_R^1. For this reason we choose to limit the term "ideal" to W_B^1, W_R^1.

There is a good interpretation for the different types of weights found in Examples 1 and 3. In Example 1 the attrition of infantry is so much greater than that of artillery that we visualize one phase of the battle ending when one side has lost all of its infantry even though both sides still have artillery left. However, at that time the starred weights do become relevant for the ensuing artillery duel.

On the other hand in Example 3 the artillery attrition is more rapid than that of infantry. Moreover, when one side runs out of artillery the remaining infantry forces will ultimately be anhilated by the surviving artillery. Hence a zero weight for infantry is not inappropriate.

Example 4 is much like Example 1 for even though $P_{B2} = P_{R2} = 0$ the larger eigenvalue λ_1^4 still gives a viable ideal weight.

5. Calculation of effectivenessmatrices and an application to Lanchester Theory. There are several possible approaches to calculation of the effectiveness matrices. Only one of these will be discussed in the present paper.

A sufficiently detailed combat simulation can be expected to produce loss matrices

(5.1) $$L_{BR} = [\ell_{BR}(i,j)] \quad , \quad L_{RB} = [\ell_{RB}(j,i)]$$

where $\ell_{BR}(i,j)$ is the number of Red weapons of class j lost by action of Blue weapons of class i, etc. Then we may define effectiveness numbers by

$$(5.2) \qquad m_{BR}(i,j) = \ell_{BR}(i,j)/u_{iB} \ , \ m_{RB}(j,i) = \ell_{RB}(j,i)/u_{jR}$$

where U_B and U_R are as in Section 1 (formulas (1.1) and (1.2)).

The u_{iB} and u_{jR} might refer either to the initial Blue and Red strengths, or to certain average strengths during the battle. The choice of an appropriate average would relate to questions not considered here; however, a simple case of such an average might be $[u_{iR(t=0)} + u_{iR(t=t_1)}]/2$ where t_1 is an arbitrary time chosen as a unit of measurement. The interval (0,t) must, of course, not exceed the battle length and should be small enough so that combat losses have not yet changed the character of the encounter.

This procedure has as its main drawbacks (1) that the validity of the results obtained depends on the simulation scenario, on the simulation model, and on the extent of sampling error, (2) that it fails to consider military appurtenances which, although affecting the combat action, do not cause attributable casualties to opposing weapon systems, and (3) that it does not take into account scale factors (i.e., it tacitly assumes that the losses are strictly proportional to the number of weapons in a class).

Effectiveness numbers calculated as above might be interpreted as estimates of the Lanchester parameters appropriate to a heterogeneous Lanchester linear system. Dare and James, in Defense Operational Analysis Establishment Memorandum M7120 have made an analysis based on this interpretation with results parallel to those given here. In Table, Appendix II to Annex L of the TATAWS III study, BAARINC Inc. has based a similar analysis on another such interpretation.

More specifically, if we have the Lanchester systems

$$(5.3) \qquad \dot{U}_B = -C_R U_R \ , \quad \dot{U}_R = -C_B U_B$$

then the (i,j) element $C_R(i,j)$ of C_R represents the effectiveness of R weapon j against B weapon i, i.e.,

$$C_R(i,j) = m_{RB}(j,i) \ .$$

Reasoning similarly for C_B we conclude that

$$(5.4) \qquad C_R = M_{RB}^T \ , \quad C_B = M_{BR}^T$$

are reasonable choices for the Lanchester coefficient matrices.

Now, differentiating equation (1.4) with respect to time we get

$$\begin{aligned}(5.5) \qquad \dot{S}(B) &= W_B^T \dot{U}_B = -W_B^T M_{RB}^T U_R \\ &= -(M_{RB} W_B)^T U_R \\ &= -(M_{RB} M_{BR} W_R / \alpha_R)^T U_R \\ &= -\frac{\lambda}{\alpha_R} W_R^T U_R = -\alpha_B W_R^T U_R \qquad (\text{since } \lambda = \alpha_R \alpha_B) \ .\end{aligned}$$

Now substituting from (1.5) this gives

$$(5.6) \qquad \dot{S}(B) = -\alpha_B S(R).$$

Similarly, differentiating (1.5) yields

$$(5.7) \qquad \dot{S}(R) = -\alpha_R S(B) \ .$$

Equations (5.6) and (5.7) are the ones obtained by Dare and James. A note of caution is appropriate here. The heterogeneous systems (5.3), (5.4) have questionable validity past the time t^* at which any component of U_R or U_B becomes zero. However, the summarizing homogeneous systems (5.6) and (5.7) will in general yield solutions $S(B)$, $S(R)$ which both remain positive far beyond t^*.

6. <u>A larger example</u>. An example of extended calculation is given below based on results obtained in a particular detailed war game. No claims are warranted concerning the representativeness of these results, which are dependent on the particular scenario, and the random statistical variation inherent in the game model used. Weapons classes for both sides were the same. They were (following some aggregation of similar type):

1. Small arms,
2. Armored personnel carriers
3. Tanks
4. Armed reconaissance vehicles
5. Anti-tank weapons
6. Mortars
7. Artillery

Red forces were in the attack, Blue in the defense.

7 Red Weapons   7 Blue Weapons

Red Effects

$$(6.1)\quad M_{RB} = \begin{bmatrix} .0145 & .0012 & .0000 & .0229 & .0004 & .0000 & .0000 \\ .0510 & .0326 & .0000 & .0638 & .0012 & .0048 & .0000 \\ .1060 & .4600 & .4540 & .4900 & .0056 & .0515 & .0000 \\ .4440 & .2220 & .0000 & .4440 & .0700 & .0000 & .0000 \\ .0000 & .1370 & .7400 & .2740 & .0137 & .0000 & .0000 \\ 6.1500 & .0000 & .0000 & .0000 & .0630 & .0740 & .0000 \\ 21.0000 & .2320 & .0750 & .2770 & .1570 & .0800 & .1960 \end{bmatrix}$$

Blue Effects

$$(6.2)\quad M_{BR} = \begin{bmatrix} .0334 & .0028 & .0000 & .0290 & .0004 & .0000 & .0000 \\ .1170 & .0940 & .0000 & .1111 & .0045 & .0000 & .0000 \\ .4770 & 2.5300 & 2.0900 & 1.8200 & .0730 & .0000 & .0000 \\ .8200 & .4730 & .0000 & .5550 & .0008 & .0000 & .0000 \\ .0000 & 2.8300 & .5000 & 3.3300 & .1860 & .1940 & .0000 \\ 12.0800 & .0000 & .0000 & .0000 & .1580 & .1502 & .0000 \\ 9.7100 & .1220 & .1000 & .1350 & .1180 & .0680 & .2590 \end{bmatrix}$$

$$(6.3)\quad P_R = \begin{bmatrix} .0194 & .0121 & .0002 & .0146 & .0001 & .0001 & .0000 \\ .1158 & .0368 & .0006 & .0445 & .0012 & .0010 & .0000 \\ 1.2978 & 1.4398 & .9517 & 1.1711 & .0448 & .0088 & .0000 \\ .4049 & .4302 & .0350 & .5171 & .0146 & .0136 & .0000 \\ .5937 & 2.0535 & 1.5534 & 1.5597 & .0574 & .0027 & .0000 \\ 1.0993 & .1955 & .0315 & .3881 & .0259 & .0233 & .0000 \\ 3.8610 & .8696 & .2548 & 1.4743 & .0801 & .0558 & .0508 \end{bmatrix}$$

$$(6.4)\quad P_B = \begin{bmatrix} .0135 & .0066 & .0003 & .0139 & .0021 & .0000 & .0000 \\ .0558 & .0285 & .0033 & .0592 & .0080 & .0005 & .0000 \\ 1.1656 & 1.4585 & 1.0029 & 2.0245 & .1433 & .1198 & .0000 \\ .2824 & .1397 & .0006 & .2956 & .0398 & .0023 & .0000 \\ 2.8689 & 1.087 & .3646 & 1.9550 & .2541 & .0537 & .0000 \\ 1.0989 & .0361 & .1169 & .3199 & .0165 & .0111 & .0000 \\ 6.0748 & .1679 & .1521 & .4432 & .0606 & .0315 & .0508 \end{bmatrix}$$

Clearly this is a reducible case with one obvious Perron eigenvalue $\lambda_2 = .0508$. Applying seven iterations we find that the other Perron eigenvalue λ_1 has the positive probability eigenvectors.

$$(6.5)\qquad Z_{1R} = \begin{bmatrix} .00052 \\ .00198 \\ .30482 \\ .03033 \\ .48015 \\ .03087 \\ .15134 \end{bmatrix}, \qquad Z_{1B} = \begin{bmatrix} .00082 \\ .00443 \\ .54613 \\ .01381 \\ .26728 \\ .10285 \\ .06480 \end{bmatrix}$$

where also

$$(6.6)\qquad \alpha_{1R} = .98947, \quad \alpha_{1B} = 1.15741$$

$$\lambda_1 = \alpha_R \alpha_B = 1.14522$$

$$W_{1R} = \alpha_{1R} Z_{1R} \quad , \quad W_{1B} = \alpha_{1B} Z_{1B} \; .$$

Since

λ_1' is much greater than λ_2, the ideal weights obtained from λ_1 may be regarded as being more significant than those obtained from λ_2 as given in (6.7) and (6.8) below.

$$(6.7) \qquad Z_{2R} = Z_{2B} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$(6.8) \qquad \alpha_{2R} = .1960 , \ \alpha_{2B} = .2590 , \ \lambda_2 = .0508 ,$$

$$W_{2R} = \alpha_{2R} Z_{2R} , \quad W_{2B} = \alpha_{2B} Z_{2B} .$$

Reproduced from best available copy.

REFERENCES

1. Georg Frobenius "Über Matrizen aus positiven Elementen", Sitzungsberichte der Kgl Preussischen Akademie der Wissenschaften zu Berlin. 1908, and Part II 1909.

2. F. R. Gantmacher, The Theory of Matrices, (2 vols.), Chelsea, 1959.

3. Oskar Perron, "Zur Theorie der Matrices" Mathematische Annalen, Vol. 64 (1907).

4. A Note on the Thor Index of Combat Effectiveness, working paper RMT-200-W10-17R, Robert M. Thrall and Associates, March 1972.

A NEW FORMULATION OF LANCHESTER COMBAT THEORY

Dr. Frank E. Grubbs
Chief Operations Research Analyst
U. S. Army Aberdeen Research & Development Center

and

Capt. John H. Shuford
Ft. Sill U. S. Army Artillery School

ABSTRACT

Lanchester's differential equations of combat are inherently deterministic in nature, although considerable effort has been devoted in recent years to introducing stochastic type treatment into the theory. Morse and Kimball (1951), for example, discussed probabilistic relations or transition probabilities for losses, and more recently Bonder (1967) introduced the idea of "variable attrition coefficients" by pointing out that the Lanchester attrition coefficients are average values or expected rates and hence by definition "imply an underlying probability distribution." Barfoot (1969) indicates that the Lanchester attrition coefficients should be estimated from the reciprocal of the mean-time-to-kill or the harmonic mean. Weiss (1957) apparently was the first to include by modelling the relative movement or the separation distances between opposing forces as an important parameter in Lanchester type theory and hence by such a hypothesis saw the need for realistic changes or variation in the attrition coefficients. Here, we advance the idea that the time-to-kill or time-to-neutralize key opposing targets would seem to be the random variable which should be treated on a probabilistic basis, and hence that the fraction of remaining combatants on each side should properly be estimated from time-to-kill probability distributions, or in other words from principles of the statistical theory of reliability and life-testing. Advantages of such treatment include the possibility that the future course of a battle may be predicted from data on casualties in the early stages of an engagement, and therefore that field commanders will thus have available information on which to base critical decisions, for example, to withdraw or to augment fighting forces in order to bring about the more desirable future courses of combat for a given mission. Also, commanders may even use analyses suggested herein independently of information on enemy losses to decide whether the course of combat is proceeding satisfactorily or according to plan by comparing data on early casualties observed in an engagement with standards which have been determined from experience or have been specified.

The new formulation is illustrated with a small scale, but informative example on an engagement between "Chief Battle Tanks" (CBT's) and "R10" type Tanks.

I. INTRODUCTION

Lanchester's differential equations of combat are really deterministic in nature, although some operations research writers on the subject (e. g. Morse and Kimball [11, 1951]) discussed probabilistic relations for numerical decreases in force size or transition probabilities, and more recently there have appeared many papers on "variable attrition coefficients". Bonder [3, 1967], for example, indicates that the attrition coefficients are "average" values or "expected rates" and therefore that such a definition "implies a probability distribution." Barfoot [1, 1969] has indicated that the attrition coefficients should be estimated from harmonic means on time and that they should therefore be determined from unity divided by the mean-time-to-kill for an individual on a side. Weiss [14, 1957] apparently first discussed Lanchester theory in the context of taking into account the movement or separation distances of forces, while Bonder [2, 1965] has made applications of Lanchester theory involving range dependent variable attrition coefficients, the values of which depend on separation distances between opposing forces, and he obtained solutions for constant relative closing velocity of forces. Very recently, Taylor [13, 1971] has generalized the range dependent attrition coefficient model to include time <u>or</u> force separation as independent variables and has established that Bonder's results [2, 1965] are special cases of his model. In spite of these various treatments of "stochastic" type combat, we believe nevertheless that there is room for a new concept or argument concerning problems of randomness and just how random variables should be treated in a realistic mathematical theory of combat. In particular, we believe it is worthwhile to hypothesize that the logical or correct random variable may be that of time, i. e., time-to-kill, and that remaining forces on each side are dependent on and tied in with time in a rather complex but random fashion. In other words, when a Blue force meets a Red force, or one stumbles upon the other, then the ensuing battle involves changing decisions on the part of commanders, the random effects of terrain, weather conditions, the selected or available weapon mixes, timely deployment and use of weapons, accidental occurrences relating to reliability and maintainability of equipment, resupply, etc., so that it is perhaps unnecessary to argue further that many conditions leading to various degrees of randomness are ever present, that the variable which should logically be treated on a probabilistic basis should be that of time-to-kill opposing targets, and therefore that other Lanchester parameters should depend in a probabilistic manner on elapsed times in battles, in particular when kills or other forms of attrition occur. As a matter of fact, if in a battle one were to tabulate the times from zero at which targets are destroyed or combatant losses occur on both sides, then he might well develop a better understanding of applied combat theory, especially in as much as general Lanchester type theory might be developed further or is really valid. But such data are usually hard to come by. Why not work the time-to-kill concept into the Lanchester type theory nevertheless to see where it might lead? This, we now proceed to do along lines similar to that covered in some detail by Shuford [12, 1971].

II. THE NEW FORMULATION

We begin with the above concept and a simple argument. Let B_o and R_o represent the initial numbers of Blue and Red combatants or targets, fighting units, etc., which are deemed appropriate as key elements or targets in an engagement, and let B and R be the numbers remaining on each side at any general time t after combat has begun. Thus, the fractions B/B_o and R/R_o each represent quantities which will vary in a random manner from unity at the start of a battle down to some fraction (or perhaps to zero), at which time the engagement ceases, or a side withdraws. Moreover, the fractions B/B_o and R/R_o clearly vary in a random manner with time, i. e. $B = B(t)$, $R = R(t)$, and indeed they are the fractions of survivors on the two sides at any time t. Therefore, it can be argued that these fractions or a function thereof could be related to various forms of probability distributions in time-to-kill. The probability distributions in time must involve meaningful physical definitions, criteria or descriptions for time-to-kill, time-to-incapacitate, time-to-failure of equipment, etc., and the parameters of such distributions should in some way describe the "fighting power" or capability of a side in the times required to kill opposing targets. To win in battle, one must kill or incapacitate before his opponent disables him. In this connection, it is well known that the two parameter Weibull distribution (actually another probability distribution of R. A. Fisher which has taken on Weibull's name!) can be used to represent a very wide variety of time-to-fail (or in this case, time-to-kill) probability distributions. Moreover, the fraction of survivors at given times in life tests of equipment is now rather widely recognized as the reliability of the equipment, so that in general such fractions could be equated to reliabilities which depend upon the random time-to-kill variables in combat. In general, for continuous distributions the reliability or fraction surviving with respect to a mission time, t_m, may be defined as the integral of an appropriate probability density function (p.d.f.) from t_m to ∞. Thus, we would have immediately the following approximations or relations for remaining fractions of Blues and Reds at any time t after the battle started:

$$B/B_o = \exp - \beta t^{\alpha} \qquad B = B(t);\ \alpha,\ \beta > 0;\ t \geq 0 \tag{1}$$

$$R/R_o = \exp - \rho t^{\delta} \qquad R = R(t);\ \rho,\ \delta > 0;\ t \geq 0 \tag{2}$$

where $\beta = \beta\{t, R, B\}$ is an "attrition" coefficient for Blue, i. e., the loss or failure rate, or scale parameter, and $\alpha = \alpha\{t, R, B\}$ a shape parameter for the time-to-kill probability distribution, which parameters represent the capability of Red forces to destroy Blue targets, Blue to protect himself, etc. In combination we might say that α and β represent in perhaps an obscure way the "total fighting power" of Red against Blue, but including also various attrition accidents which occur to Blue in battle. Similar arguments apply to (2). By the notation $\beta = \beta\{t, R, B\}$, for example, we mean here that β is the parameter (constant) of a distribution which is statistically estimable from the probabilistic relation between the remaining Blues and Reds with time.

We might well derive (1) and (2), of course, somewhat formally from the consideration that $B/B_o = B(t)/B_o$ is the fraction of Blues remaining at time t, or it can be referred to as the chance that a Blue combatant, tank target, or fighting unit, etc., will survive to time t, and hence that $(B_o - B)/B_o$ is the chance of a Blue combatant being lost by time t. Thus, we may hypothesize that $(B_o - B)/B_o$ is the cumulative chance of kill for Blues within the random time t and that furthermore the time derivative of this quantity can be equated to a probability density function of time-to-kill. In summary, we say, for example, that

$$\frac{dF}{dt} = \frac{1}{B_o} \frac{d(B_o - B)}{dt} = \alpha\beta t^{\alpha-1} \exp - \beta t^{\alpha} \tag{3}$$

where the left-hand side is the fractional rate of losses for Blue and the righthand side is the two-parameter Weibull p. d. f. for time-to-kill Blue targets. Integrating (3), we obtain

$$B = B_o \exp - \beta t^{\alpha} \tag{4}$$

The Weibull p.d.f. has been used because of its inherent generality in describing accurately various shapes of time-to-kill distributions occurring in combat.

Also, we could argue that since $(B_o - B)/B_o$ is the fraction of losses for Blue, then the conditional failure rate for Blue, given survival to some time t, may be described somewhat generally in the form

$$\frac{B'(t)/B_o}{B(t)/B_o} = - \alpha\beta t^{\alpha-1} \tag{5}$$

where the right-hand side depends on the time of battle. That is to say, the conditional failure rate of Blues may vary with some power of time, possessing the generality of an increasing, constant or decreasing kill rate. Hence, we get immediately that

$$\ln[B(t)/B_o] = - \beta t^{\alpha}$$

or as before

$$B(t)/B_o = e^{-\beta t^{\alpha}}$$

Now the fractions of survivors, or the "reliabilities" given by (1) and (2), as we have indicated, can really encompass a wide range of probability distributions on time for combat type engagements. In fact, the two-parameter Weibull p. d. f. given by

$$f(t) = \alpha\beta t^{\alpha-1} \exp - \beta t^{\alpha} \qquad (6)$$

is somewhat of a natural choice for it can, by proper selection of the shape and scale parameters, α and β, vary from the sub-exponential, to the exponential (in which case $\alpha = 1$, and the conditional failure or kill rate is constant and equal to β) to the super-exponential models of time-to-kill. Indeed, various combinations of α and β even include the normal or Gaussian p. d. f., as well as skew, platykurtic or leptokurtic type probability distributions. We can therefore through the use of the Weibull model or theory equate the "random" fractions of Blue and Red survivors to any of a wide variety of realistic probability distributions for remaining lives, which in some way will depend on the "fighting powers" or combat capability of the opposing sides. If the probability of survival of Blue forces consistently exceeds that of Red, then Blue obviously has the advantage in an engagement.

We next consider the problem of estimating the parameters of the time-to-kill probability distributions.

III. PARAMETER ESTIMATION FOR EXPONENTIAL TIME-TO-KILL DISTRIBUTIONS

In case combat losses as a function of time take on a purely exponential form of decay (i. e. constant conditional failure or kill rate), as may sometimes occur, then $\alpha = \delta = 1$, and (1) and (2) become simply

$$B/B_o = \exp - \beta t \qquad \beta > 0,\ t \geq 0 \qquad (7)$$

$$R/R_o = \exp - \rho t \qquad \rho > 0,\ t \geq 0 \qquad (8)$$

In this case, a tabulation of times at which targets are killed on both sides would give the information needed for widely known, best estimates of parameters β and ρ (Epstein and Sobel [6, 1953].) Thus, if we have B_o initial Blues, and the times-to-kill Blue targets are in the natural ascending order

$$t_1 \leq t_2 \leq \cdots \leq t_r \leq \cdots \leq t_{B_o}$$

where we may truncate the battle at the time of the rth Blue casualty, or base estimates of β on the first r Blue casualties, for example, then the maximum likelihood, minimum variance, best unbiased estimator of the true unknown mean-time-to-kill, i. e. $\theta = 1/\beta$, is (Epstein and Sobel [6, 1953])

$$\text{Estimate } 1/\beta = \hat{\theta} = [\sum_{i=1}^{r} t_i + (B_o - r)\, t_r]/r \qquad (9)$$

This is simply the total of kill times for their Blues and survival times for the remaining $B_o - r$ Blues to the rth kill, or the "total time on test", as it is called, divided by the number of actual kills.

Alternatively, and from statistical considerations of unbiasedness, we may equate the observed fraction of kills for Blue up to time t_i to the quantity

$$(B_o - B_i)/B_o \approx i/(B_o + 1) \qquad i = 1, 2, \ldots, B_o \tag{10}$$

or that is

$$(B_i/B_o) \approx (B_o - i + 1)/(B_o + 1) \tag{11}$$

and linearize (7) by taking logarithms to the base e. We get in this case that

$$\ln\,[(B_o + 1)/(B_o - i + 1)] = \beta t_i \tag{12}$$

for $i = 1, 2, \ldots, B_o$ and the increasing observed values of t_i. Thus, β may be estimated by least squares, for example, by using (12). Formula (12) in fact may be particularly desirable to justify the exponential time-to-kill hypothesis, or may be used routinely for large scale or untruncated battles, although (9) is the universal estimate of the parameter for an exponential time-to-kill distribution.

IV. PARAMETER ESTIMATION FOR THE WEIBULL MODEL

Estimates of the scale and shape parameters for the two-parameter Weibull models (1) and (2) are available from literature on the statistical theory of reliability, either for truncated or complete samples. Because of space limitations, we cannot cover in any detail the overall problem of estimation in Weibull theory, as the volume of literature on the general subject is great indeed. In fact, Weibull estimation and confidence limit theory is almost a branch of statistics in its own right. Consequently, we refer the reader, for example, to the paper of Cohen [4, 1965] for maximum likelihood estimation, and the papers of Mann[9, 10; 1966, 1967] which in the example below use the linear invariant statistics. Also, a recent paper, "Statistical Inference from Censored Weibull Samples" by B. R. Billman, C. E. Antle and L. J. Bain, submitted to TECHNOMETRICS would be of considerable use.

It could be argued that there is an advantage in the routine use of the Weibull model for our purposes here, for given any data one may proceed to estimate the unknown scale and shape parameters by the above references, or others in the literature, thereby arriving at the appropriate form of the actual time-to-kill distributions. It may be desirable in most cases to program the estimation of parameters on a computer, along with other descriptions of the battle as discussed below.

From (1) and (2), we note that these Weibull forms may easily be linearized by taking logarithms to the base e twice, obtaining

$$\ln\ln B_o/B = \ln\beta + \alpha\ln t \tag{13}$$

This linear equation would be particularly useful for estimating the scale and shape parameters from graphical considerations, for sample sizes beyond the tables of Mann (1966, 1967), or routinely for rather large-scale or complete simulations, etc. In fact, one may have to adapt the methods of estimation to the particular problem at hand, and also program computations for obtaining both the parameter estimates and confidence limits on the remaining fractions of survivors for a given mission time.

V. ADDITIONAL CONSIDERATIONS

For other possible models based on the idea raised here, we could, of course, argue that losses on each side must depend not only on the value of the parameters, α, β, ρ, and δ, or that is on the attrition distributions which occur as a result of weapon mixes, tactics, terrain, etc., but rather that losses must be related directly to opposing numbers of combatants, units, etc., as in the Lanchester Square Law. Thus, we might set up models such as

$$\frac{d}{dt}\left[(B_o - B)/B_o\right] = f(t)\left[R/R_o\right] \tag{14}$$

$$\frac{d}{dt}\left[(R_o - R)/R_o\right] = g(t)\left[B/B_o\right] \tag{15}$$

where $f(t)$ and $g(t)$ are again time-to-kill probability density functions. Actually, Taylor [13, 1971] and others have studied the equivalent of a somewhat related case, $f(t)/g(t) = k$, a constant, but rather our argument here is that the remaining fractions of survivors on the two sides are precisely by definition the <u>reliabilities</u> for time-to-kill probability distributions, i. e.

$$B(t)/B_o = 1 - F(t) = \exp - \beta t^{\alpha}; \qquad F(t) = \int_o^t f(x)\,dx$$

Thus, there would seem to be some advantage in the herein suggested treatment of Lanchester type combat theory, for we could simulate a field exercise or fight a battle on a computer with the mixes of weapons, tactics, etc., we desire, and then stop the simulation at some appropriate number of targets lost on a side, which would lead to sufficiently accurate estimates of the scale and shape parameters. Once these estimates of the Weibull parameters are available from time-to-kill data, then the straight-forward deterministic solutions of (1) and (2) give the predicted characteristics of the battle at any desired times, or, as indicated below, we may also derive confidence statements on the remaining fractions of Blues and Reds for any (mission) time of the battle. Consequently, it is clear that we can draw upon available statistical theory of reliability and life-testing to save time in simulations, effectiveness studies, systems evaluations and the like,

since our approach (although not so limited) may be to analyze the first, relatively few, times to casualties, without the necessity of drawn out computer simulations or very lengthy war games, in order to see from early stages how well our new equipment, strategies, etc., may actually work in a hypothesized combat situation.

To emphasize, therefore, the time variable would seem to be of some central importance in an overall, realistic treatment of Lanchester type combat theory, and the time-to-kill targets, or time-to-neutralize an opposing force, etc., would appear to be of a critical character in describing outcomes of engagements. Thus, more emphasis should be placed on tracing a battle, simulation, war game or the like in the time variable, and in particular the times at which casualties occur. We therefore have a straight-forward, natural and economical way to proceed with studies of various mixes of weapons, hypothesized optimum tactics, or other considerations, and possibly large amounts of time or costs otherwise might well be saved. We reemphasize that to determine the variable rates of attrition as a function of time it is important to estimate the shape characteristics of the time-to-kill distributions for Blue and Red, and compute values therefrom. Indeed, such distributions for various engagements might be synthesized to predict combined arms capabilities.

Of course, we have treated the numbers of Blues and Reds here as continuous variables, so that opposing numbers should be rather large generally for such treatment, although the inherent discreteness could be taken care of mathematically if needed. On the other hand, we are sampling hypothesized populations or probability distributions, and many times relatively large numbers of combatants will be involved or required anyway in simulations to infer general or precise battle outcomes for complex situations.

The chance that a Blue survives to time t or beyond is clearly $\exp - \beta t^{\alpha}$, and the chance that a Blue is put out of action by time t is therefore $1 - \exp - \beta t^{\alpha}$. The conditional probability that a Blue survives to time t and then is put out of action during the next small increment of time Δt, is given by $\alpha\beta t^{\alpha-1} \cdot \exp - \beta t^{\alpha} \Delta t / \exp - \beta t^{\alpha} = \alpha\beta t^{\alpha-1} \Delta t$. Similar quantities also hold for the Red side. With such basic probabilities, therefore, one can compute for a given time the chances that various numbers of Blues and Reds are lost (or are surviving) out of the initial numbers of Blues and Reds by using Binomial probabilities. Alternatively, the conditional probabilities may be used for a variety of calculations such as a Blue or Red winning a duel in a short period of time, or other probabilities of interest could be computed.

For illustrative purposes, we now give an example on application of the above methodology. Although the engagement discussed in the example is rather small scale and involves only tanks, it should nevertheless indicate how the methodology might apply to large-scale games or even to heterogeneous weapon mix studies as indicated alternatively in Section VII. We do not delve into criteria for victory or defeat, probability of winning or losing, and the like, for such outcomes may well depend on unidentifiable or intangible factors. On the other hand, we do indicate a method for estimating and placing confidence limits on remaining fighting forces, which may well affect the outcome of battle.

VI. EXAMPLE

In a study of the effectiveness of anti-tank missiles as the main armament of tanks, it was decided to simulate a "typical" engagement in Western Europe for a certain version of the CBT (Chief Battle Tank) versus the R 10. One of the main purposes of the simulation was to determine whether missiles could successfully engage opposing tanks at longer ranges than guns and hence obtain an early advantage in killing enemy tanks, thereby neutralizing the enemy tank force and obtaining a given objective on schedule. In particular, a mission time of about 90 minutes was suggested for accomplishment of the objective. In a valley, twenty R 10's were in position near the bottom of an inclining ground area leading up to a town of key importance in the hills of the general battle zone. The R 10's were initially defiladed in position and hence not easily in view of the friendly task force of 20 CBT's approaching them. At about 2500 meters, however, the R 10's opened fire on the approaching CTB's, but the latter were out of range for very accurate fire from the R 10's. As a result, and as the battle proceeded, the first tank knocked out was a R 10 by the approaching CBT's at four minutes after the engagement had started. In eight minutes, one CBT had come within range of the R 10's and was killed. In summary, five R 10's were knocked out at 4, 9, 15, 23, and 40 minutes elapsed time from the beginning of the engagement. On the other hand, three CBT's were killed at 8, 13 and 24 minutes and later at 60 minutes another CBT was finally knocked out. Between the period 40-60 minutes, it was thought that some other R10's had been put out of action, but a heavy fog had set in, making such determination uncertain, and the battle was stopped just before night. With these data on times-to-kill targets on each side, and assuming no major changes in the commanders' tactics, resupply, etc., what could be said about the progress and outcome of such a battle in general had it continued, assuming the above represents valid sampling for a population of such engagements?

We assume that the time-to-kill distributions for tank targets on each side follow two-parameter Weibull probability distributions because of the wide variety of possible shapes for fitting such data and we proceed to estimate the parameters, so that an appropriate fit can be obtained which would describe the probable remainder of such an engagement. For quickness and convenience, we will use the tables of Mann[9, 1966] to estimate α, β, δ and ρ, although other methods of estimation could be used, for example the maximum likelihood estimates of Cohen[4, 1965], or that of Billman, Antle and Bain referred to above. In order to use the estimates of Mann[9, 1966, p. 47], i. e. the linear invariant statistics, it is convenient to tabulate the computations as follows:

CBT Data ($B_o = 20$)				R 10 Data ($R_o = 20$)			
Times-to-kill		Mann's Coefficients or Weights		Times-to-kill		Mann's Coefficients or Weights	
t_i	$\ln t_i$	A_i	C_i	t_i	$\ln t_i$	A_i	C_i
8	2.079	-.408	-.244	4	1.386	-.273	-.193
13	2.565	-.386	-.239	9	2.198	-.259	-.191
24	3.178	-.346	-.223	15	2.708	-.234	-.181
60	4.094	2.141	.706	23	3.136	-.200	-.166
				40	4.689	1.965	.732

$\Sigma A_i \ln t_i = 5.827 = \hat{u}$

$\Sigma C_i \ln t_i = 1.061 = 1/\hat{\alpha}$

Thus, $\hat{\alpha} \approx 1/1.061 \approx .943$

and

$\hat{\beta} = e^{-\hat{\alpha}\hat{u}} \approx 1/244 \approx .0041$

$\Sigma A_i \ln t_i = 5.040 = \hat{u}$

$\Sigma C_i \ln t_i = 1.002 = 1/\hat{\delta}$

Thus, $\hat{\delta} \approx 1/1.002 \approx .998$

and

$\hat{\rho} = e^{-\hat{\delta}\hat{u}} \approx 1/154 = .0065$

From the above, we note that since the estimates of shape parameters α and δ are each practically one, then exponential time-to-kill distributions may be used to describe the battle, i. e., the losses on each side. In fact, the estimated true mean-time-to-kill a CBT for the population would be 244 minutes, whereas the mean-time-to-kill a R 10 is estimated to be only 154 minutes. Put another way, and since the exponential failure distribution involves a constant conditional failure rate at any time t, then the failure or kill rate for CBT's is predicted as .0041 per minute, and that for R10's is .0065 per minute.

Since the single parameter negative exponential distribution seems to be a suitable hypothesis from the above estimates of shape parameters (≈ 1) for the small numbers of kills, we could well estimate the scale parameters, β and ρ, i. e. the conditional failure rates from formula (9). We have in fact

$$\text{Est } 1/\beta = \frac{\sum_{i=1}^{r} t_i + (B_o - r)t_r}{r} \qquad (r = \text{number kills})$$

$$= \frac{105 + (16)(60)}{4} = 266 \qquad (\text{vs. } 244)$$

$$\text{Est } 1/\rho = \frac{91 + (15)(40)}{5} = 138 \qquad (\text{vs. } 154)$$

so that the agreement is surprisingly good in this case.

An interesting and important feature of the methodology suggested herein is that we may easily place confidence limits on the fractions of survivors for each side. For example, for the assumption of an exponential distribution, it is known (Epstein and Sobel [6, 1953]) that

$$2r\hat{\theta}/\theta = \chi^2(2r) \tag{16}$$

where $\theta = 1/\beta$ or $1/\rho$, $\hat{\theta} = 1/\hat{\beta}$ or $1/\hat{\rho}$, and $\chi^2(2r)$ = Chi-Square with 2r degrees of freedom. That is to say, $2r\beta/\hat{\beta}$, and $2r\rho/\hat{\rho}$, are each distributed in probability as the well-known Chi-Square and hence since the true unknown fraction of Blue survivors is $e^{-\beta t}$, and that for Red is $e^{-\rho t}$, we may determine confidence limits for the true fractions of survivors as follows. We start with

$$\Pr[\chi^2_\alpha(2r) \leq \chi^2(2r) = 2r\ \hat{\theta}/\theta \leq \chi^2_{1-\alpha}\ (2r)] = 1 - 2\alpha \tag{17}$$

where χ^2_α is the lower α probability level and $\chi^2_{1-\alpha}$ the upper α probability level of the Chi-Square distribution for 2r degrees of freedom. Hence, for a "mission" time t_m we can convert the above probability statement to

$$\Pr[t_m\ \chi^2_\alpha\ (2r)/2r\ \hat{\theta} \leq t_m/\theta < t_m\ \chi^2_{1-\alpha}\ (2r)/2r\hat{\theta}] \tag{18}$$

$$= \Pr[\exp - \{t_m\ \chi^2_{1-\alpha}\ (2r)/2r\hat{\theta}\} \leq e^{-t_m/\theta} \leq \exp - \{t_m\ \chi^2_\alpha\ (2r)/2r\hat{\theta}\}] \tag{19}$$

$$= 1 - 2\alpha$$

But

$$\exp - \beta t_m = B/B_o \qquad \text{and} \qquad \exp - \rho t_m = R/R_o$$

for any mission time t_m and thus we have lower and upper confidence limits on the fractions of Blue and Red survivors. Thus, had the tank battle gone to 1 1/2 hours (90 minutes), we could state for the assumption of an exponential distribution that

$$\Pr[B/B_o \geq \exp - \{t_m\ \hat{\beta}\ \chi^2_{1-\alpha}\ (2r)/2r\}] = 1 - \alpha$$

or

$$\Pr[B/B_o \geq \exp - \{(90)(1/266)\ \chi^2_{.95}\ (8)/8\} = .52] = .95$$

or in other words we state with 95% confidence that at least 52% (10.4) of the CBT's would survive after 90 minutes of such a battle. On the other hand, we could only say that at least 30.4% (6.1) of the R10's would survive after 90 minutes, again with 95% confidence.

With two-sided confidence limits based on $\chi^2_{.025}(8) = 2.18$, $\chi^2_{.975}(8) = 17.53$ $\chi^2_{.025}(10) = 3.25$ and $\chi^2_{.975}(10) = 20.48$, we could state with 95% confidence that at 90 minutes the fraction of surviving CBT's will be between .48 and .91, whereas for the same confidence level the fraction of surviving R 10's will lie between .26 and .81. Of course, the widths of the confidence intervals depend markedly on the number of kills, the conditional failure rate, the mission time and the confidence level, and in this illustration we are dealing with rather sparse data from a rather limited engagement to infer very precise statements about the general population.

For the mission time of 90 minutes, the estimated fractions for point estimates of surviving CBT's and R 10's would be respectively exp - (90/266) ≈ .71 and exp - (90)/138 ≈ .52. Should more precise information be desired, then the simulation could be carried further, repeated, or the problem enlarged in consonance with the importance of the decision to be made.

We remark that similar confidence limits could be estimated for the Weibull distributions (i. e. when $\alpha \neq 1$ and $\delta \neq 1$) in (1) and (2). See, for example, Johns and Lieberman [7, 1966].

VII. PROBABILITY ANALYSIS FOR MIXED WEAPONS OR COMBINED ARMS STUDIES

Although the above theory may be useful in combined arms studies, i. e. the times-to-kill may be analyzed as data from a single population (at least to some extent) no matter what weapon fires at what target, it may be desirable or in some cases necessary, to take into account the particular weapon types used against particular targets, especially in so far as capability of weapons is concerned. In other words, we have at hand a technique which can be employed to compare the overall kill potential of one class of weapons versus that of another against common targets, etc. Also, it could be very informative to keep types and classes of weapons separate in a simulation or war game, at least up to some stage of battle, and then determine whether the various kill distributions might be combined into a composite or single distribution which would describe the overall combined arms effects.

For analyses of this type, we let β_{ij} and α_{ij} represent respectively the scale and shape parameters which identify the capability of Red's jth weapon type to destroy Blue's ith target in a time-to-kill distribution. Likewise, we let ρ_{ij} and δ_{ij} represent the scale and shape parameters which describe the capability of Blue's ith weapon type to destroy Red's jth target in a combat situation. With these definitions, it is clear that we could proceed as follows for a probability analysis.

Since

$$\exp - \beta_{ij} t^{\alpha_{ij}}$$

is equal to the chance that the ith Blue target survives the jth Red weapon within time t, then clearly

With two-sided confidence limits based on $\chi^2_{.025}$ (8) = 2.18, $\chi^2_{.975}$ (8) = 17.53 $\chi^2_{.025}$ (10) = 3.25 and $\chi^2_{.975}$ (10) = 20.48, we could state with 95% confidence that at 90 minutes the fraction of surviving CBT's will be between .48 and .91, whereas for the same confidence level the fraction of surviving R 10's will lie between .26 and .81. Of course, the widths of the confidence intervals depend markedly on the number of kills, the conditional failure rate, the mission time and the confidence level, and in this illustration we are dealing with rather sparse data from a rather limited engagement to infer very precise statements about the general population.

For the mission time of 90 minutes, the estimated fractions for point estimates of surviving CBT's and R 10's would be respectively exp - (90/266) ≈ .71 and exp - (90)/138 ≈ .52. Should more precise information be desired, then the simulation could be carried further, repeated, or the problem enlarged in consonance with the importance of the decision to be made.

We remark that similar confidence limits could be estimated for the Weibull distributions (i. e. when $\alpha \neq 1$ and $\delta \neq 1$) in (1) and (2). See, for example, Johns and Lieberman [7, 1966].

VII. PROBABILITY ANALYSIS FOR MIXED WEAPONS OR COMBINED ARMS STUDIES

Although the above theory may be useful in combined arms studies, i. e. the times-to-kill may be analyzed as data from a single population (at least to some extent) no matter what weapon fires at what target, it may be desirable or in some cases necessary, to take into account the particular weapon types used against particular targets, especially in so far as capability of weapons is concerned. In other words, we have at hand a technique which can be employed to compare the overall kill potential of one class of weapons versus that of another against common targets, etc. Also, it could be very informative to keep types and classes of weapons separate in a simulation or war game, at least up to some stage of battle, and then determine whether the various kill distributions might be combined into a composite or single distribution which would describe the overall combined arms effects.

For analyses of this type, we let β_{ij} and α_{ij} represent respectively the scale and shape parameters which identify the capability of Red's jth weapon type to destroy Blue's ith target in a time-to-kill distribution. Likewise, we let ρ_{ij} and δ_{ij} represent the scale and shape parameters which describe the capability of Blue's ith weapon type to destroy Red's jth target in a combat situation. With these definitions, it is clear that we could proceed as follows for a probability analysis.

Since

$$\exp - \beta_{ij} t^{\alpha_{ij}}$$

is equal to the chance that the ith Blue target survives the jth Red weapon within time t, then clearly

$$\prod_{j=1}^{n} \exp - \beta_{ij} t^{\alpha_{ij}} = \exp - \sum_{j=1}^{n} \beta_{ij} t^{\alpha_{ij}} \tag{20}$$

is the chance that the ith Blue target survives all of the n Red weapons which can possibly destroy it.

Similarly, the chance that the jth Red target will survive all of the m Blue weapons which could engage it by time t is

$$\prod_{i=1}^{m} \exp - \rho_{ij} t^{\delta_{ij}} = \exp - \sum_{i=1}^{m} \rho_{ij} t^{\delta_{ij}} \tag{21}$$

By subtracting in turn the quantities (20) and (21) from unity we get respectively the chance that the ith Blue target is put out of action by at least one of Red's weapons and the chance that the jth Red target is put out of action by at least one of Blue's weapons which has such capability. Straightforward enumeration leads to probabilities that various combinations (or all) of Blue's or Red's targets would be put out of action (or would survive) by some given time in a battle.

VIII. FINAL REFLECTIONS

As mentioned earlier, it is difficult under ordinary circumstances to obtain times at which casualties occur in actual battles, and especially such data for the opposing side. Nevertheless, in realistic simulations of battles or computer games, etc., one can indeed acquire the needed data and hence have at hand information to judge the probable future outcomes of engagements using the methodology suggested herein. Also, data obtained in a natural manner on the friendly side, with no such information at all on enemy casualties may be of considerable importance. For example, the field Army carries as part of its equipment nowadays some computers, so that if Blue were in a battle and had been allocated a certain time, say three hours, to accomplish an objective, then computations could be made in the field, and during the battle, to estimate from the Blue casualties occurring, say, during the first 30, 45 or 60 minutes of battle, just what the shape of the appropriate Weibull p. d. f. might be, and hence predict the remaining blue survivors at the mission time of three hours. (We remark in this connection that truncating a simulation or battle at some predetermined fixed time as compared to that of a fixed number of casualties would lead to somewhat different methods of estimation.) If this estimated fraction of survivors is expected or is satisfactory, then Blue proceeds, but otherwise higher headquarters would be so advised and hence have important information on

which to base any decision to withdraw, throw additional units into the battle, etc. Furthermore, standard values of the Weibull parameters, β and α, might be developed from experience, and hence computed casualties as a function of time might be compared with observed rates in a simulation or actual battle to determine whether requirements are satisfactorily met, or various alternative actions should be taken accordingly by commanders.

Finally, other forms of probability distributions could, of course, be fitted to observed time-to-kill data on targets in a battle or simulation, although it is believed that the two-parameter Weibull model suggested here would represent a single form of distribution which should be sufficient for many battle situations of interest.

REFERENCES

1. C. B. BARFOOT, "The Attrition-Rate Coefficient, Some Comments on Seth Bonder's Paper and a Suggested Alternative Method," *Operations Research 17*, 888-894 (1969).

2. SETH BONDER, "A Theory for Weapon System Analysis," *Proc. U. S. Army Operations Research Symposium*, 111-128 (1965).

3. SETH BONDER, "The Lanchester Attrition-Rate Coefficient," *Operations Research, 15*, 221-232 (1967).

4. A. C. COHEN, "Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and on Censored Samples", *Technometrics, 7*, 579-588 (1965).

5. J. H. ENGEL, "A Verification of the Lanchester's Law", *Operations Research 4*, 148-151 (1956).

6. B. EPSTEIN AND M. SOBEL, "Life-Testing", *Journal of the American Statistical Association 48*, 486-502 (1953).

7. M. V. JOHNS AND G. J. LIEBERMAN, "An Exact Asymptotically Efficient Confidence Bound for Reliability in the Case of the Weibull Distribution," *Technometrics 8*, 135-174, (1966).

8. F. W. LANCHESTER, *Aircraft in Warfare, the Dawn of the Fourth Arm*, Constable, London, 1916.

9. N. R. MANN, "Tables for Obtaining the Best Linear Invariant Estimates of Parameters of the Weibull Distribution." *Rocketdyne Research Report 66-9*, Rocketdyne, Canoga Park, California. (1966).

10. N. R. MANN, "Tables for Obtaining the Best Linear Invariant Estimates of Parameters of the Weibull Distribution", *Technometrics 9*, 629-645 (1967).

11. PHILIP M. MORSE AND GEORGE E. KIMBALL, *Methods of Operations Research*, John Wiley, New York 1951.

12. JOHN H. SHUFORD, "A New Probability Model for Lanchester's Equations of Combat", *Masters Thesis submitted to the George Washington University*, December 1971.

13. JAMES G. TAYLOR, "A Note on the Solution to Lanchester Type Equations with Variable Coefficients", *Operations Research 19*, 709-712 (1971).

14. H. K.WEISS, "Lanchester-Type Models of Warfare," *Proc. First International Conference on Operations Research*, pp 82-89, December 1957.

A NOTE ON THE THOR INDEX OF COMBAT EFFECTIVENESS

J. R. Thompson and R. M. Thrall

Let there be given two sides in a combat - the Blue and the Red - with force strengths $u = (u_1, u_2, \ldots, u_m)$ and $v = (v_1, v_2, \ldots, v_n)$. Consider as possible measures of their respective strengths at time t the linear functions

$$(1) \qquad \begin{cases} B(u(t),v(t)) = \sum_{j=1}^{m} a_j u_j(t), \text{ and} \\ R(u(t),v(t)) = \sum_{i=1}^{n} b_i v_i(t) \end{cases}$$

where the coefficients a_j and b_i are non-negative quantities which may be functions of t. Suitable selection of these coefficients is an important problem to which some attention is given below.

The THOR index at time t is defined as

$$(2) \qquad T(u(t),v(t)) = \frac{B(u(t),v(t))}{R(u(t),v(t))}$$

and provides a measure of the effectiveness of the Blue side relative to that of the Red side.

In the important case $m = n = 1$ (called the lumped case) we propose as one possibility the definitions:

$$(3) \qquad \begin{cases} a = \sqrt{\frac{dv}{dt} \Big/ \frac{du}{dt}} \\ b = \sqrt{\frac{du}{dt} \Big/ \frac{dv}{dt}} \end{cases},$$

giving

$$(4) \qquad T(u(t),v(t)) = \frac{\left(\frac{dv}{dt}\right)}{\left(\frac{du}{dt}\right)} \frac{u}{v},$$

Preceding page blank

all quantities are to be evaluated at time t. Here a geometric interpretation for T is obtained by writing the equations of lines passing through $u(t)$ and $v(t)$ with slopes $\frac{du}{dt}$ and $\frac{dv}{dt}$ respectively:

$$(5) \qquad \begin{cases} u^*(\tau) = u(t) + \frac{du}{dt}(\tau - t) \\ \\ v^*(\tau) = v(t) + \frac{dv}{dt}(\tau - t) \, . \end{cases}$$

Let t_u and t_v be the time values giving $u^*(t_u) = 0$ and $v^*(t_v) = 0$, respectively. Then

$$(6) \qquad T(u(t),v(t)) = \frac{(t_u - t)}{(t_v - t)} \, .$$

Let us consider how (4) applies to a simple hypothetical example. Suppose that 5 Blue regiments engage 4 Red regiments in combat. At the end of the fifth day, Blue intelligence learns that 2 Red regiments remain. At the same time, the strength of the Blue force is 4 giments. We wish to assess the THOR index.

Analysis # 1. If we assume no governing combat law, we estimate the derivative directly from the data, obtaining

$$(7) \qquad T(u(5),v(5)) = \frac{\left(-\frac{2}{5}\right)}{\left(-\frac{1}{5}\right)} \frac{4}{2} = 4 \, .$$

Analysis # 2. If we assume a Lanchester Law of the form

$$(8) \qquad \begin{cases} \frac{du}{dt} = -c_1 u^q v \\ \\ \frac{dv}{dt} = -c_2 u v^p \end{cases}$$

where $0 \leq q \leq 1$ and $0 \leq p \leq 1$, then

(9) $$\frac{u^{2-q}}{2-q} = \frac{c_1}{c_2}\frac{v^{2-p}}{2-p} + \left(\frac{u_0^{2-q}}{2-q} - \frac{c_1}{c_2}\frac{v_0^{2-p}}{2-p}\right)$$

and

(10) $$T(u(t),v(t)) = \frac{c_1}{c_2}\frac{u^{2-q}}{v^{2-p}} .$$

Now, Blue wins if and only if

(11) $$\left(\frac{2-p}{2-q}\right)\frac{c_2}{c_1}\frac{u^{2-q}}{v^{2-p}} > 1$$

or, equivalently:

(12) $$T(u(t),v(t)) > \frac{2-q}{2-p} .$$

Below we show Table I demonstrating $\frac{c_2}{c_1}$ and T values for various values of p and q using the data from the hypothetical combat example.

TABLE I

	$p=q=0$	$p=q=1$	$p=1,\ q=0$
$\frac{c_2}{c_1}$	$1\frac{1}{3}$	2	$\frac{4}{9}$
T	$5\frac{1}{3}$	4	$3\frac{5}{9}$

It is interesting to note that the T values for each of the model assumptions agree closely with each other and with the earlier value in which no model assumptions were made.

The generalization of (3) to the multiforce case presents some problems since an ineffective subforce with good protection would receive a higher coefficient than a subforce which fights effectively but sustains casualties.

We have not yet completed our study of the coefficients for the multi-

component case, but will present two further possibilities.

One is to employ a Delphi procedure (i.e., to use the pooled judgement of military experts), and the other is an analytic determination of the effect at time t on each component of one side of a change in each initial component for the other side.

In both cases we let

$$(13)\qquad \begin{cases} a_{ji}(t) = \text{measure of a change in } v_i(t) \text{ induced by a change} \\ \qquad\qquad \text{in } u_j(0) \\ \\ b_{ij}(t) = \text{measure of a change in } u_j(t) \text{ induced by a change} \\ \qquad\qquad \text{in } v_i(0)\ . \end{cases}$$

Then, we would define

$$(14)\qquad \begin{cases} a_j(t) = \Sigma_i\ w_{ji}(t)a_{ji}(t) \\ \\ b_i(t) = \Sigma_j\ z_{ij}(t)b_{ij}(t) \end{cases}$$

where the w_{ji} and z_{ij} are weights to be assigned. The simplest case would be $nw_{ji} = mw_{ij} = 1$ for all i and j. A more attractive possibility is

$$(15)\qquad \begin{cases} w_{ji} = w_i = \Sigma_\ell\ b_{i\ell}/\Sigma_{k\ell}b_{k\ell} \\ \\ z_{ij} = z_j = \Sigma_\ell\ a_{j\ell}/\Sigma_{k\ell}a_{k\ell}\ . \end{cases}$$

In the Delphi case we could let

$$a_{ji} = (a'_{ji} + a''_{ji})/2$$

where a'_{ji}, a''_{ji} are respectively, the averages of the experts' answers to the two questions (1) "What would be the expected incremental destruction by

the end of the battle to Red subforce i if Blue subforce j is increased by one unit" and (2)"What would be the likely decrease in destruction by the end of the battle to Red subforce i if Blue subforce j is decreased by one unit" . Here the phrase "end of the battle" means the earlier of the times at which the two battles under consideration end. The definition of the b_{ij} is strictly analogous. The values of a_{ji} and b_{ij} thus obtained are not functions of time, but they do depend on the "initial" force vectors $u(0),v(0)$ and hence would need to be determined for a broad spectrum of "initial" forces.

In the second (analytic) approach we wish to define a partial derivative of v_j with respect to $u_i(0)$. If we assume that each component of u and v depends on time and the starting force vectors $u(0),v(0)$ we may write

$$
(16) \qquad \begin{cases} u(t) = h(u(0),v(0),t) \\ v(t) = k(u(0),v(0),t) \end{cases}
$$

Let $\mathcal{E}_p^q$ denote the p-th basis vector in q-space. Then we define

$$
(17) \qquad \begin{cases} a_{ji}(t) = \dfrac{\partial v_j(t)}{\partial u_i(0)} = \lim\limits_{\tau\to 0} \dfrac{k(u(0) + \tau\mathcal{E}_i^m,v(0),t) - k(u(0),v(0),t)}{\tau} \\ b_{ij}(t) = \dfrac{\partial u_i(t)}{\partial v_j(0)} = \lim\limits_{\tau\to 0} \dfrac{h(u(0),v(0) + \tau\mathcal{E}_j^n,t) - h(u(0),v(0),t)}{\tau} \end{cases}
$$

These functions a_{ji}, b_{ij} have the common property that they all vanish for $t = 0$ (since no change in one initial force vector can have any effect on any component of the opponent's <u>initial</u> force vector). This implies that $T(0)$ is formally indeterminate. However, $\lim\limits_{t\to 0} T(t)$ may very well exist and be calculable, especially if u and v are defined via differential equations (e.g., a heterogeneous Lanchester model).

We postpone consideration of the general case and illustrate (17) for the situation described by equations (8) above. We get

$$(18)\qquad \begin{cases} u = \sigma(v) = [u_0^{2-q} + \frac{2-q}{2-p} \cdot \frac{c_1}{c_2} (v^{2-p} - v_0^{2-p})]^{1/(2-q)} \\ \\ v = \rho(u) = [v_0^{2-p} + \frac{2-p}{2-q} \cdot \frac{c_2}{c_1} (u^{2-q} - u_0^{2-q})]^{1/(2-p)}. \end{cases}$$

Now, from (8) we get

$$(19)\qquad \begin{cases} \dot{v} = \frac{dv}{dt} = -c_2\ \sigma(v)v^p \\ \\ \dot{u} = \frac{du}{dt} = -c_1\ \rho(u)u^q \end{cases}$$

Separating variables and integrating we get

$$(20)\qquad \int_{v_0}^{v} \frac{dv^*}{v^{*p}\sigma v^*} + c_2 t = 0\ .$$

If we differentiate (20) with respect to u_0 (using Leibnitz' rule for differentiating a definite integral) we obtain (cf (17))

$$(21)\qquad \frac{\frac{\partial v}{\partial u_0}}{v^p\sigma(v)} - \int_{v_0}^{v} \frac{u_0^{1-q}\ dv^*}{v^{*p}\sigma(v^*)^{3-q}} = 0,$$

and hence

$$(22)\qquad a(t) = \frac{\partial v}{\partial u_0} = v^p\sigma(v)u_0^{1-q} \int_{v_0}^{v} \frac{dv^*}{v^{*p}\sigma(v^*)^{3-q}}\ .$$

Now

$$(23)\qquad \lim_{t\to 0} \frac{a(t)}{t} = (v_0^p\ \sigma(v_0)u_0^{1-q})\ \frac{\dot{v}(0)}{v_0^p\ u_0^{3-q}} = \frac{\dot{v}(0)}{u_0}\ .$$

Similarly

$$(24)\qquad b(t) = \frac{\partial v}{\partial u_0}\ u^q\rho(u)v_0^{1-p} \int_{u_0}^{u} \frac{du^*}{u^{*q}\rho(u^*)^{3-p}}$$

and

$$\lim_{t\to 0} \frac{b(t)}{t} = \frac{\dot{u}(0)}{v_0} . \tag{25}$$

Now

$$T(t) = \frac{a(t)u(t)}{b(t)v(t)} = \frac{(a(t)/t)u(t)}{(b(t)/t)v(t)} \tag{26}$$

so that

$$T^* = \lim_{t\to 0} T(t) = \frac{\dot{v}(0)}{\dot{u}(0)} = \frac{c_2\ u_0^{1-q}}{c_1\ v_0^{1-p}} . \tag{27}$$

This index T^* has the disadvantage that when $p = q = 1$ it is independent of the initial force structure.

We may prefer, by analogy with (3) to redefine $a(t),b(t)$ as the square roots of the partial derivatives $\frac{\partial v}{\partial u_0}$, $\frac{\partial u}{\partial v_0}$. This leads to the new index

$$T^{**} = \left(\frac{\dot{v}(0)}{u_0}\ \frac{v_0}{\dot{u}(0)}\right)^{\frac{1}{2}} \cdot \frac{u_0}{v_0} = \sqrt{\frac{c_2 u_0^{2-q}}{c_1 v_0^{2-p}}} \tag{28}$$

which is the square root of the index given in (10). Using this new index T^{**} the bottom row of Table I would read $\frac{\sqrt{48}}{3}$, 2, $\frac{\sqrt{32}}{3}$ or approximately 2.3, 2, 1.9.

Another problem to which we hope to address ourselves is a consideration of cases in which it is not assumed that the battle proceeds until annihilation of one of the sides is achieved. A more reasonable analysis might be based on the following formulation.

Let $z(t) = (u_1(t),\ldots,u_m(t),\ v_1(t),\ldots,v_n(t),t)$. Then R^{m+n+1} is partitioned into the four regions:

$$C_B = \{z(t) \mid \text{Blue (but not Red) abandons the combat}\}$$

$$C_R = \{z(t) \mid \text{Red (but not Blue) abandons the combat}\}$$

$$C_{BR} = \{z(t) \mid \text{Both Red and Blue abandon the combat}\}$$

$$C = \{z(t) \mid \text{combat continues}\} .$$

This would suggest as a combat index

$$D = \text{Prob}(z \text{ will cross from } C \text{ to } C_R) + \tfrac{1}{2}\,\text{Prob}(z \text{ will cross from } C \text{ to } C_{BR})$$

An alternative index is suggested by the combat example given earlier in this note. Suppose that in that example, Blue is willing to fight only until its force has been reduced to three regiments, but Red is willing to fight until its force has been annihilated. Then

$$(t_u - t) = \frac{-1}{\frac{du}{dt}}$$

$$(t_v - t) = \frac{-2}{\frac{dv}{dt}} .$$

Using the numerical estimates for the values of $\frac{du}{dt}$ and $\frac{dv}{dt}$ obtained (without model assumptions), we have

$$T' = \frac{t_u - t}{t_v - t} = \frac{-1}{\left(-\frac{1}{5}\right)} \Big/ \frac{-2}{\left(-\frac{2}{5}\right)} = 1 .$$

Usint T' as an index, we would say that Blue and Red are now on a parity as to who will hold the field at the end of the engagement. Thus, in a sense, willingness to accept high casualties may increase a side's combat effectiveness.

DECISION RISK ANALYSIS FOR RESEARCH AND DEVELOPMENT

Dr. John D. Hwang and Dr. C. Richard Shumway*
U.S. Army Air Mobility R&D Laboratory
Ames Research Center, Moffett Field, California

SUMMARY

This paper examines decision risk analysis for the management of technology-oriented research and development. Two stages are proposed: (1) A local decision risk analysis to insure research objectives are accomplished; and, (2) A global analysis to allocate limited resources optimally. An operational definition of risk for military R&D is derived in terms of the classical statistical risk definition. This definition follows a natural hierarchy from research projects to project outcomes, contributions to technologies, and military utility. Data are collected from both scientists/engineers and managers without compromising their subjective evaluations. More significantly, this paper points out the very natural tie between risk analysis and R&D resource allocation.

I. INTRODUCTION

In July 1971, the Department of Defense Directive (DODD) 5000.1, entitled "Acquisition of Major Defense Systems", was officially released. Subsequently, several key defense officials described the roles and implementation plans to such a directive (Defense Management Journal, 1971). For a research and development laboratory which is primarily involved in basic research, exploratory development, and portions of advanced development through the demonstration of technology, the objective is to establish a strong and usable technology base and to transform ideas and technology into defense systems which fulfill defense needs. Major guidelines in DODD 5000.1 which are directly relevant to an R&D laboratory include the following:

a. Establishment of a strong technology base.

b. Needs and requirements matched with technology.

c. Trade-offs made on cost, time, and capability.

d. Optimal resource allocation ensured.

e. Risk assessment carried out.

R&D management is, of course, constantly hampered by many constraints, imposed primarily by the limited available resources. On the one hand,

*Assistant Professor of Economics, North Carolina State University at Raleigh.

the R&D community is directed to establish a strong technology base. On the other, the R&D efforts are restricted by limited available resources. Consequently, R&D management is faced with two major problems: (1) For each research project, will the research objectives be accomplished by the project layout?; and, (2) How are limited resources to be allocated among all R&D efforts? In view of the great emphasis on the subject "decision risk analysis" throughout the Department of Defense since 1969, this paper has the objective to show that decision risk analysis assists the decision-makers in R&D management in handling the above problems. After some preliminary remarks on decision risk analysis, we exhibit a methodology of decision risk analysis which attempts to answer the first question. In Chapter III, risk analysis is properly interfaced with the resource allocation problem which addresses the second problem. We conclude this paper with some suggested areas of research.

Since the inception of the subject of decision risk analysis, a major problem has been the lack of a generally accepted set of definitions; and, as a result, the methodology to conduct such an analysis is muddled. Intuitively, it is clear what risk means; for instance, a model was constructed by Aerospace Industries Association (1969) which portrays the conversion of unknowns to knowns relative to the progression in the materiel acquisition process. Two types of unknowns are highlighted which affect the three key dimensions of cost, time, and performance of any materiel system or process: known-unknowns, and unknown-unknowns or "unk-unks". Attempts were made to resort to allied areas such as systems analysis/operations research, statistical decision theory, utility theory, etc., to define decision risk analysis. With very minimum DOD guidance (Packard, 1970) that the intent of risk analysis is to reduce program risk by formal risk assessment of technical problems, system and hardware proofing, and risk avoidance trade-offs, risk analysis was then blessed with a "catch-all" definition:

"A disciplined process, essential to program decision making, involving the application of a broad class of qualitative and quantitative techniques in analyzing, reducing, and assessing uncertainties associated with the realization of cost, time, and performance goals of large-scale military projects."

In spite of the situation, decision risk analysis evolved rapidly. Risk has been introduced as a fourth dimension in addition to cost, time and performance, and risk analysis has been envisaged as "systems analysis of risk" (Hwang and Arnett, 1970). Some prototype risk analyses were carried out and are found in Hwang (1971) and throughout the U.S. Army Materiel Command.

The U.S. Air Force Academy conducted a review of the subject and proposed a list of candidate definitions as follows (1971):

Risk - Probability that a planned event will not be attained within constraints (cost, schedule, performance) by following a specified course of action.

Uncertainty - Incomplete knowledge.

Risk Assessment - A comprehensive and structured process for estimating the risk associated with a particular alternative course of action; also the product of such a process.

Risk Management - The generation of alternative courses of action for reducing risk.

Risk Analysis - The process of combining the risk assessment with risk management in an iterative cycle; also the product of such a process.

We propose to extend the definition to include the following:

Decision Risk Analysis - The formulation of risk analysis highlighting alternative courses of actions and consequences for purposes of management decision-making.

This added definition is significant. If the analysis and results are not easily understood by the decision-makers, the reluctance of decision-makers to trust such an analysis creates credibility gaps, and all is lost.

Methodologies and operational definitions have been developed in numerous associated disciplines and are applicable to decision risk analysis. Utility theory has been most helpful to quantify preferences. Gamble or lottery techniques have been proposed to develop a priori probability laws over the states. Additionally, the "Delphi" technique offers some possibilities for magnitude estimation and collection of opinions. Quantification of contractor risk has been proposed. An iso-risk contour generation scheme was also developed (Hwang, 1970). Special simulation techniques that compute time or cost variations include Program Evaluation and Review Technique (PERT), Critical Path Method (CPM), Graphical Evaluation and Review Technique (GERT), to name a few.

II. LOCAL DECISION RISK ANALYSIS

In the consideration of a research project, a program is structured with objective, approach, schedule, and resources needed. A basic question that confronts the decision-maker is whether or not the research objectives will be achieved by following a specified course of action with the planned schedule and programmed resources. If not, what can the decision-maker do to manage more effectively? He is in need of knowing possible outcomes and consequences, to anticipate possible failures, and to have planned for alternative courses of action. This kind of prior information can be generated in part from a decision risk analysis with an assessment of potential technical problems, consequences of failure, judgment as to efforts needed to resolve problems, and impacts on schedule and total cost. A generalized methodology has been proposed in the form of a closed decision analysis feedback cycle (Howard, 1966) consisting of deterministic-probabalistic-informational phases. In the following, let us consider an

operational model involving a specific case of a V/STOL tilt rotor research aircraft project (Hwang et al., 1972) jointly developed by the U.S. Army and NASA.

The approach to the operational model subscribes to the standard decision tree analysis (Raiffa, 1968) and consists of the following steps:

1. Aggregate all R&D efforts into major phases.

2. For each phase, identify potential problems, and assess probability of occurrences.

3. For each problem, evaluate consequences of failure.

4. Enumerate means to resolve problems and attach probabilities of success to each.

5. Estimate impacts on schedule and on cost.

6. Fold back for expected values (or apply Monte-Carlo simulation for outcome distributions).

7. Check sensitivity.

The V/STOL tilt-rotor aircraft under analysis features two large diameter rotors mounted on tiltable wing-tip nacelles. The rotors provide hover comparable to a helicopter but can tilt by rotating the nacelles 90° so that the rotors operate as conventional propellers for cruising. The objective of the project is to develop a proof-of-concept demonstrator and research vehicle, and to show that current technology is adequate for the development of a useful commercial or military V/STOL tilt-rotor aircraft. R&D efforts are aggregated into the following major phases:

1. Engineering design and tooling
2. Bench tests
3. Ground test
4. Wind tunnel test
5. Flight test

Data were collected from technical personnel and revealed some thirty potential problems, some with high probabilities of occurrence. However, the high probability of success in resolving all potential problems substantiates the fact that there is a very sound technology base. The analysis provided the project management with a probability of program success, possible cost growth, and schedule delay for potential technical problems, and a basis for contingency fund, as well as alternative courses of actions should the potential problems occur.

We refer the interested readers to the study itself and do not elaborate this local risk analysis in this paper, for the significance of this paper rests in the global level.

III. A GLOBAL RESOURCE ALLOCATION MODEL

A large number of specific models have been developed for resource allocation among R&D activities, many of which are directed to Defense R&D. Through various decision-rule and optimization approaches, they purport to analyze certain data in a prespecified manner and suggest an appropriate allocation of available resources. These models vary greatly in scope and procedure. Some terminate with a rank ordering of R&D projects; others attempt to allocate funds and/or manpower optimally. Some are deterministic models; others incorporate stochastic elements. Most are static; a few are dynamic. A number are constrained in focus to economic evaluations; others are unrestricted as to type of variables considered. A few attempt to derive the optimal size of the R&D budget, but most assume the overall funding for R&D to be an upper restraint to the decision problem.

The publication of these models is widely scattered, but there are three basic references which briefly describe a number of them and provide extensive bibliographic references to many more: Baker and Pound (1964); Cetron, Martino, and Roepcke (1967); and Baker and Freeland (1972).

The following model is an innovative approach to the R&D resource allocation problem developed at the Army Air Mobility R&D Laboratory. It is an adaptation of a basic model introduced by A.B. Nutt (1965), includes a modified value function, also a piecewise-linear solution capability, and formally incorporates risk.

Let us proceed first with a discussion of global risk. The classical definition of risk as found in most standard statistics texts (DeGroot, 1970) is:

$$\rho(P,d) = \int_{\Omega} L(w,d)\, d\, P(w)$$

Where $\rho(P,d)$ is the expected loss or risk, of loss $L(w,d)$ for any decision d and outcome w with a probability distribution P on the outcome space Ω. It is fairly standard to specify the loss as:

$$L(w,d) = -\, U\, [\sigma(w,d)]$$

where $\sigma(w,d)$ is the reward for each decision d and each outcome w, and U is a utility function on the set of rewards. Since we are interested in choosing a decision d which minimizes the risk, Bayes risk $\rho^*(P)$ is a possible candidate which is defined as the greatest lower bound, "minimum", for the risks $\rho(P,d)$ for all decisions d, i.e.,

$$\rho^*(P) = \text{g.l.b. } \rho(P,d)$$

For a discrete outcome space, the risk definition is as follows:

$$\rho(P,d) = -\sum_{w_i \in \Omega} U\, [\sigma(w_i,d)]\, p(w_i)$$

where $p(w)$ is the probability density function corresponding to $P(w)$.

Let us interpret the above definition for a research and development laboratory management. The simplified diagram below depicts the basic hierarchy from R&D projects to the missions.

Research Projects	Probability of Outcome Occurrence	Project Outcome	Outcome Contribution	Military Utility
	p_{11}	w_{11}	$\sigma(w_{11},d)$	U_{11}
X_1	p_{12}	w_{12}	$\sigma(w_{12},d)$	U_{12}
	p_{13}	w_{13}	$\sigma(w_{13},d)$	U_{13}
X_2	p_{21}	w_{21}	$\sigma(w_{21},d)$	U_{21}
	p_{22}	w_{22}	$\sigma(w_{22},d)$	U_{22}

. . .

Assume that local decision risk analyses have been conducted on individual projects. These analyses have resulted in a plan of work at alternative funding options which minimizes the risk of failing to achieve the project objective(s). One possible resource allocation decision d includes the subordinate decision to fund a given project X_1 at a particular level. Corresponding to this decision is a set of probabilities (p_{11}, p_{12}, p_{13}) of project outcomes (w_{11}, w_{12}, w_{13}). The anticipated contribution of each possible outcome to technology objectives is measured by $\sigma(w_{11},d)$, $\sigma(w_{12},d)$, and $\sigma(w_{13},d)$. The overall military utility of achievement of the technology objectives is reflected by the U_{11}, U_{12}, and U_{13}.

To further clarify the diagram, we consider the case of air mobility research and development. The basic missions of air mobility are intelligence, fire power, mobility, command, communication and control, and logistical support. R&D projects are carried out in the development of technology areas of aerodynamics, structures, propulsion, militarization, support, avionics, weaponization, and system synthesis. A typical project such as the V/STOL tilt-rotor research aircraft can have a number of possible outcomes: successful tilt-rotor aircraft with desired flying qualities, complete failure, or intermediate stages whereby certain significant knowledge is acquired not previously known or predicted. Each outcome reveals information which contributes to the technologies of aerodynamics, structure, and propulsion. The information also has military utility relative to the air mobility missions.

Associated with each project, we have tasks or work units with alternative funding levels, scientific man-years, and test facility requirements attached. Aggregate resource requirements and subset requirements must fall within specified limits:

1. Budget Constraints -

a. Combined task or work unit funding must not exceed the total R&D budget.

$$\sum_{jk} C_{jk} X_{jk} \leq B$$

b. Reprogramming authority between program elements is restricted, after funds are appropriated.

$$B_{e^-} \leq \sum_{jk} C_{jke} X_{jk} \leq B_{e^+}$$

2. Manpower restraints by manpower type introduce an upper limit on the availability of certain groups of scientists, and a lower limit to assure minimum capability if reduction-in-force takes place.

$$M_t^- \leq \sum_{jk} M_{jkt} X_{jk} \leq M_t^+$$

3. Test facility time requirements by type of facility may not exceed their availability.

$$\sum_{jk} F_{jk\,f} X_{jk} \leq F_f^+$$

The following symbols are defined:

e - program element
f - facility type
j - work unit identification
k - funding alternative
t - manpower type

B - budget
C - cost
F - facility
M - manpower
X - work unit

To obtain solution, the global minimization model is couched in a piecewise linear framework. That is, if two funding options are specified with corresponding contributions and utilities, a linear combination of the funding options is assumed possible with a corresponding linear combination of contributions and utilities. Therefore, if multiple funding options are specified, the computer algorithm is designed to select any option or any point on a linear segment between sequential options. This assumption permits solution of the problem using "separable programming" or, if only one resource is constraining, "network analysis". The IBM separable programming package for the 360-series computers is readily available and has been used satisfactorily, as has a network program prepared by Baker, Jarvis and Unger (1971).

The data collection phase consists of various distinct levels. The laboratory technical staff and executive committee identify laboratory

objectives, describe the most important products of each objective desired within given time periods, and weigh the objectives relative to military utility, or Army needs (U). Scientists and engineers identify the possible outcomes (w) of project efforts and probability of occurrence (p) at each alternative funding level (C) based partially upon a local risk analysis. They also specify manpower (M) and test facility (F) requirements by type at each funding option. They address the outcome contributions to technologies as well ( ). It is important to note that under this scheme, data are collected from both scientists/ engineers, and managers without compromising their subjective evaluations.

IV. CONCLUSION

In summary, we propose that decision risk analysis for the management of technology-oriented research and development consists of two major parts. A local decision risk analysis is conducted on each project to insure that research objectives are accomplished. A global decision risk analysis is carried out to allocate limited resources among all R&D efforts optimally.

The global decision risk analysis model is one which minimizes risk in terms of contribution to technology and military utility, subject to resource constraints on budget, manpower, and facilities. Most significantly, the risk model is derived from the classical risk definition and is totally consistent with resource allocation models which maximize expected military utility.

The global resource allocation model has been experimentally applied to partial research programs at the U.S. Army Air Mobility R&D Laboratory. A more thorough experiment preceding its implementation Laboratory-wide will be conducted in the near future. A consistent rating scheme designed to reflect the Laboratory objectives and military utility should prove to be a challenge to any analyst.

Decision risk analysis is a valuable management decision-making tool which systematically and simultaneously evaluates data which could not otherwise be adequately analyzed. It is neither intended to solve all management problems, nor designed to be an automated decision-maker. It does encourage participative management throughout the organization.

REFERENCES

1. Aerospace Industries Association, "Essential Technical Steps and Related Uncertainties in DOD Weapon Systems Development", (Phases I, II, III, IV), Washington, D.C., December 1970.

2. Baker, N. R. and J.R. Freeland, "Recent Advances in R&D Value Measurement and Project Selection Methods", 41st National Meeting of ORSA, New Orleans, April 1972.

3. Baker, N. R., J. J. Jarvis, and V. E. Unger, "Documentation for OPTYAG: A Research and Development Resource Allocation Program", U.S. Army Air Mobility R&D Laboratory, Ames Research Center, Moffett Field, California, August 1971.

4. Baker, N. R. and W. H. Pound, "R and D Project Selection: Where We Stand", IEEE Transactions on Engineering Management, December 1964.

5. Baker, N. R., C. R. Shumway, et al, "An R&D Project Selection and Budget Allocation Model for Large, Decentralized Research Organizations", 12th American Meeting of TIMS, Detriot, Michigan, September 1971.

6. Cetron, M. J., J. Martino, and L. A. Roepcke, "The Selection of R&D Program Content - Survey of Quantitative Methods", IEEE Transactions on Engineering Management, March 1967.

7. Defense Management Journal, Office of Assistant Secretary of Defense (Installations and Logistics), Pentagon, Washington, D.C., Fall, 1971.

8. DeGroot, Morris H., Optimal Statistical Decisions, McGraw-Hill, 1970.

9. Department of Defense Directive 5000.1, "Acquisition of Major Defense Systems", July 13, 1971.

10. "Final Report of the USAF Academy Risk Analysis Study Team," U.S. Air Force Academy, Colorado, 1 August 1971.

11. Howard, Ronald A., "Decision Analysis: Applied Decision Theory", 4th International Conference on Operations Research, Boston, 1966.

12. Hwang, John D., "Analysis of Risk for the Materiel Acquisition Process - Part I: Fundamentals", Report SY-R6-70, U.S. Army Weapons Command, Rock Island, Illinois, November 1970.

13. Hwang, John D., "Risk Analysis Versus Systems Analysis for the Materiel Acquisition Process", Proceedings - Tenth U.S. Army Operations Research Symposium, Army Research Office, Durham, N.C., May 1971.

14. Hwang, John D. and J. Arnett, "Risk Analysis", Defense Industry Bulletin, December, 1970.

15. Hwang, John D., et al, "Risk Analysis of a V/STOL Tilt Rotor Research Aircraft", U.S. Army Air Mobility R&D Laboratory and NASA V/STOL Projects Office, Ames Research Center, Moffett Field, California, March 1972.

16. Nutt, N. B., "An Approach to Research and Development Effectiveness", IEEE Transaction on Engineering Management, September 1965.

17. Packard, David, "Weapon Systems Acquisition Policy", Army Research and Development News Magazine, July-August 1970.

18. Raiffa, H., Decision Analysis, Addison-Wesley, 1968.

ACKNOWLEDGMENT

The authors gratefully acknowledge LTC Vernon Speicher, HQ, USAMC, for the identification of possible discrepancies on "risk analysis" and "decision risk analysis" between this paper and Army Materiel Command guidance. Readers are referred to the official definitions and distinctions in the letter from Deputy Commanding General, AMCPA-SA, "Decision Risk Analysis in AMC", HQ, U. S. Army Materiel Command, Washington, D. C., 3 June 1971.

RISK ANALYSIS IN MILITARY R&D PROJECTS

MAJ A. K. Varnell
US Army Logistics Doctrine, Systems & Readiness Agency
New Cumberland, PA 17070

CPT D. McGrath
US Army Combat Developments Command
Fort Monmouth, NJ 07703

Mr. D. Salvano
Mr. W. Holmes
US Army Electronics Command
Fort Monmouth, NJ 07703

INTRODUCTION

The need for an organized, logically appealing and easily applied approach to risk analysis in major Research and Development projects has been recognized for some time. Scientific management techniques have been employed by the Department of the Army (DA) and, more specifically, by the Army Materiel Command (AMC) at all phases of the development, procurement and production of weapons systems and materiel. DA Pamphlet 11-25, The Management Process for Development of Army Systems, and AMC Regulation 11-27, Conceptual Model - Life Cycle Management of US Army Materiel and AMC Official Milestones, provide these detailed procedures and constitute a logical framework for application of a formal risk analysis. Indeed these regulations have assumed, explicitly, that a formal assessment of risk is undertaken. However, the role of risk analysis has been understated and relegated to the whims of intuition and subjective opinion - with disastrous yet highly predictable consequences. Confronted with significant budget reductions and increasing Congressional scepticism, the military decision-maker must have a proven technique for assessing risk. This tool should provide quantitative evaluations, at pre-defined milestones, of the relative risk associated with pursuing various alternatives. Lacking this his decisions will remain vulnerable to attack and without authority. This paper has been directed toward meeting this need. A viable technique for formal assessment of risk, to be used in conjunction with the Army Life Cycle Model, is developed and presented.

BACKGROUND

The problem of decision-making under uncertainty is not a new one. It has been addressed by Pascal, Bernoulli and

Bayes in the 17th and 18th centuries. However, applications to real world problems were never seriously considered until relatively recent times. The work of Morgenstern and Von Neuman in this area provided a reinforcing influence to the awakening interest in scientific approaches to management.

Making decisions while operating in the uncertain environment is the very essence of the management problem. Only those problems involving uncertainty are of interest, and they surely constitute the large majority of the decision-making in the Department of the Army. It is in this area that the great challenges lie; and, accordingly, where the most lucrative benefits are to be derived. For "the question is not whether uncertainty exists, but rather in determining the nature and magnitude of the uncertainty." (1)

Recently the literature has significantly expanded with the publication of various books, articles and dissertations on the broad subject of decision-making under uncertainty. Increasingly, references to and explicit mention of the term "risk analysis" have appeared. In its most general sense risk analysis may be defined as "the broad class of techniques for analyzing, quantifying and reducing the large set of uncertain events that are inevitably associated with the realization of time, cost and performance goals of large scale military projects." (2) Obviously the approach need not be limited solely to military projects; application to large civilian projects is apparent. A more restrictive definition was that given recently by Mr. David Packard, Under Secretary of Defense, "Risk assessment is a careful assessment of the technical problems involved and a judgement as to how much effort is likely to be necessary in finding a solution that is practicable."

As noted previously risk analysis has received increasing attention lately, both in literature and by decision makers at the highest levels. No approach has yet been developed within the Department of the Army to quantitatively assess and compare the risk associated with pursuing various alternatives in large Research and Development projects. This paper is intended to fill this void by incorporating the concepts of many papers, on the general subject of decision-making under uncertainty, into a procedure based upon the Army Life Cycle Model.

The Army Life Cycle Model provided the obvious base on which to apply a formal risk analysis. It is a rigorous,

well-documented and publicized management process. It closely resembles the approach used in civilian industry. Within the model, certain key decision points were identified as being especially critical. At these milestones a series of pertinent questions were developed keyed to the nature and magnitude of the many uncertainties associated with the project at that decision point. A scoring procedure was devised to reduce all questions to a common measure of risk using weight and worth functions (to be explained later). The system flexibility is made possible by allowing the widest latitude in modifying and changing the questions asked, and applying personal utility values to the scoring. It is this feature, together with the inherent simplicity and ease of application in providing quantitative results that gives this approach utility.

DISCUSSION

The decision-maker in large scale military Research and Development projects is generally confronted with four types of uncertainty. These are technology, cost, schedule and enemy threat. This last uncertainty is normally regarded as being beyond the control of the decision-maker and the most pessimistic evaluation is usually taken as given.

It is generally accepted that the central problem in most Research and Development projects is technology. Many experts in the field, if such persons do in fact exist, will attribute 60 to 90 percent of overall project risk to technology, especially during the early phase of the project, concept formulation* (3). However, it must be immediately recognized that the three uncertainties of immediate concern, cost, schedule, and technology are interdependent. One may be reduced by increasing either or both of the remaining, although the relationship is not linear and will vary over different projects.

A considerable effort has been expended in analyzing and estimating cost uncertainties in large projects. Costs overruns for many Air Force projects were classified according to degree of technological advancement (4). A positive correlation with initial technical uncertainty was obtained. However,

*While this theory has won generally universal acceptance Perry takes issue. He states that technological uncertainty may not be the significant factor in injecting risk, it might be unwieldy and poorly functioning organizations.

even within a given technological advance class, substantial cost variations exist. A Rand Corporation study has shown that cost increases of up to 1,000 percent are not uncommon and have even exceeded 2,200 percent. This can probably be explained by the fact that military R&D efforts are inherently riskier than civilian efforts, which tend to be short term, safe and keyed toward modest advances in the state-of-the-art. Additionally, optimistic bidding on military R&D projects appears to be the nature of the competitive environment.

The project manager has some means available for combating technical uncertainty. The foremost of these would be pursuing as many feasible alternatives as practicable. Hence he can keep his options open and remain flexible. Another approach would require the use of prototypes in the development stage. Many unanticipated problems can be quickly uncovered if tested, proven prototypes are necessary prior to entering production. Finally, the project manager can suspend further operations and insist upon renewed efforts in basic and applied research if technical problems continue during development.

A manager's success is measured, to a large degree, by his facility for rapid, accurate determination of the nature and extent of these uncertainties. He must gauge their impact upon his program and immediately take measures to eliminate, or at least reduce them. Early recognition of tenuous situations is vital. Thus, any approach to risk appraisal must be geared to early identification and isolation of high risk situations. It cannot be valuable otherwise.

Previously risk assessment was a purely subjective judgment by the decision-maker or his staff analyst. Perhaps the lack of a formal, logically conceived approach was more responsible for this situation than lack of awareness of the value of risk analysis. A recent study has shown that decisions made by business executives do not necessarily emanate from a rigorous logical process (5). By failing to write out the facts, array them in logical fashion, and examine the conclusions drawn, these experienced decision-makers arrived at irrational decisions. Thus it can be stated that any successful approach to risk analysis must provide a formal, written, logically consistent technique capable of early application in the project, when correct decisions reap the most lucrative rewards.

METHODOLOGY

The central problem is one of risk assessment. However, prior to resolving this some other, more immediate, tasks must be accomplished. As previously mentioned a systematic procedure has already been delineated for management of large military R&D projects. With this structure, the Life Cycle Model, the actions and interactions of all applicable agencies, Department of Defense, Department of the Army, Combat Developments Command, Army Materiel Command are specified from the earliest stage of Concept Formulation through procurement and, literally, to the retirement of the particular equipment from the inventory. Certain critical milestones in the evolution of the model are evident. These are points where rigorous evaluation of the project status must be performed to determine whether continuing efforts, and hence deeper commitments shall be made. At these milestones there is implicit recognition that a judicious application of risk analysis is required. Early and successful application of risk analysis should potentially yield the greatest benefit. Consequently, the milestones chosen for formal application of risk analysis are:

1. Preparation of Proposed Material Need Technical Plan (AMC Activity #16).

2. Advanced Development Plan (AMC Activity AD2).

3. Proposed System Development Plan (AMC Activity #28C).

4. Contract Definition IPR/SSE (AMC Activity 53).

The first three were chosen during the Concept Formulation phase, the last occurs during Contract Definition. It should be understood that these milestones were chosen as the junctures where AMC interaction with the other major commands was required. Essentially, this approach was developed for use at the AMC Commodity Command level; yet, as the possible application extends to civilian projects so also does it apply to other Department of the Army commands. For each of these decision points a series of questions was prepared. Divided into three sections, each group was devised to expose and isolate as much information as possible on the uncertainties associated with technology, cost and schedule. The decision maker can opt to remove or add questions as desired.

Armed with these questions, a means to reduce their answers to a common measure of risk is required. Figure 1 shows the

probable impact of each type of uncertainty upon risk during project development. While these values are suggested for use in this approach, others can be inputted if considered more realistic in any specific application. Each series of questions is assigned its contributing proportion to overall project risk. In addition each question is assessed a weight, a numerical score ranging from 1 to 10. Weight is defined as a numerical evaluation of the impact of a particular question upon determination of risk in relation to all other questions in the same categories, i.e., technology, cost or schedule.

UNCERTAINTY	PERCENT OF PROJECT RISK		
	CONCEPT FORMULATION	CONTRACT DEFINITION	DEVELOPMENT AND PRODUCTION
PERFORMANCE	80 - 90%	60 - 70%	30 - 40%
COST	5 - 10%	15 - 20%	30 - 50%
SCHEDULE	5 - 10%	15 - 20%	30 - 40%

Figure 1

When weights have been assigned to all questions, each individual weight is multiplied by 100 and divided by the sum of weights of the section it is in. This figure when multiplied by the section percentage of risk contribution results in the modified weight of the question. Now the impact of the answer upon total risk is expressed in percentage form. For example, Figure 2 shows three questions for each of the major sections. Percentage contribution toward risk is shown as 70, 20 and 10 percent for technical, cost and schedule uncertainties respectively. Each individual question has a modified weight and the sum of all modified weights total to 100 percent.

The approach to be used here makes use of a technique widely applied throughout Department of Defense and attributed to Alain Enthoven (6). This is the BOP method in which an uncertainty is treated by providing a range of possible values from Optimistic through the Best single answer to the most Pessimistic. If the answer to a question is unfavorable, that is, tending to diminish the probability of successful

project completion, then the total modified weight of that question is added to the previous total of risk points assessed. Unfavorable responses may be affirmative or negative and should be intuitively obvious. No weight points are assigned upon receiving a favorable reply. Each question is posed until all have been answered. An overall risk determination, anywhere from 0 to 100, has been obtained and this procedure can be repeated for any number of alternatives. Thus various alternatives can be compared and ranked and the basis for a logical choice exists.

%	UNCERTAINTY SECTION	QUESTION WEIGHT	RELATIVE WEIGHT	MODIFIED WEIGHT
70	Technology	8	$8 \times \frac{100}{20} = 40$	28
		8	$8 \times \frac{100}{20} = 40$	28
		4	$4 \times \frac{100}{20} = 20$	14
		20		
20	Cost	9	$9 \times \frac{100}{20} = 45$	09
		7	$7 \times \frac{100}{20} = 35$	07
		4	$4 \times \frac{100}{20} = 20$	04
		20		
10	Schedule	6	$6 \times \frac{100}{18} = 33$	03
		9	$9 \times \frac{100}{18} = 50$	05
		3	$3 \times \frac{100}{18} = 17$	02
		18		100%
100%				

Figure 2

So far, this approach gives one estimate of risk for each alternative. The Best-Optimistic-Pessimistic range is obtained

by employing a worth concept. For some questions which either are not entirely relevant or for which no single answer exists or for which confidence in the response is lacking, a series of rules is employed. To obtain an Optimistic estimate of risk, a favorable response is taken to the specific question during that pass and no weight points are added. Conversely the Pessimistic estimate demands that during that pass an unfavorable response be taken and weight points added accordingly. To compute the best estimate of project risk an additional modifier, Worth, must be defined. Worth is a measure of the relevance of a particular question and/or the confidence the decision-maker has in the accuracy of the response. It is measured on a scale of 0.5 to 1.0. Questions considered totally irrelevant should be disregarded, however, questions which generate a wide diversity of answers should never be cast out. The worth rating is multiplied by the modified weight assessed the question. If an unfavorable response is in question then the risk points added are the multiple of Worth and Modified Weight. However, for favorable responses the risk points added are the product of (1.0 - Worth) and Modified Weight. In this manner the Best estimate of project risk is compiled. Figure 3 provides an example of BOP scoring for a simple, nine-question program. Obviously for those questions, to which the decision-maker is assured of the relevance and accuracy of the response, the Worth assessed is 1.0. In these instances, the Pessimistic, Optimistic, and Best estimates would receive the same value.

It must be recognized that the estimate of risk obtained has certain limitations. While the results should be reasonably accurate and provide a meaningful basis for decision-making, estimates should always be used for comparison with other alternatives at the same milestone or with estimates obtained at succeeding milestones. Failure to consider all the possible, pertinent questions could distort the results obtained. In addition, bias can also be introduced by prejudicial weighting of the questions. Another potential hazard could arise if invalid conclusions are drawn from a comparison.

EVALUATION OF RISK ESTIMATES

The utility of this approach, as explained earlier, to the decision-maker is largely dependent upon the analysis of the risk estimates obtained for the various alternatives. In selecting his preference he has unfortunately, considerable latitude. There is no single, most powerful test available except for those rare and unimportant cases where

QUESTION	MODIFIED WEIGHT	WORTH	FAVORABLE* RESPONSE	PESSIMISTIC ESTIMATE	OPTIMISTIC ESTIMATE	BEST ESTIMATE
1	15	1.0		15	15	15
2	5	1.0	X	0	0	0
3	8	0.6	X	8	0	3.2
4	12	1.0		12	12	12
5	10	0.7		10	0	7
6	10	1.0		10	10	10
7	14	1.0	X	0	0	0
8	6	1.0		6	6	6
9	20	0.8	X	20	0	4
	100%			81%	43%	57.2%

*Indicates that answer to question increases the probability of successfully meeting the project goals.

Figure 3

dominance exists. Figure 4 shows the results of an analysis where Alternative A is clearly dominant over the other strategies.

		ALTERNATIVE				
		A	B	C	D	E
RISK ESTIMATE	Pessimistic	50	65	80	75	70
	Best	40	55	50	65	60
	Optimistic	30	50	30	60	50

Figure 4

The situation more likely to occur might be that shown in Figure 5. Here no one alternative exhibits dominance over all others.

		ALTERNATIVE				
		A	B	C	D	E
RISK ESTIMATE	Pessimistic	75	80	65	55	85
	Best	40	45	60	50	80
	Optmistic	30	10	55	45	75

Figure 5

However, Alternative E can be eliminated from further consideration. It is clearly less desirable than any of the other alternatives. Figure 6 lists various criteria for selection of the most desirable alternative and ranks the choices.

CRITERIA	RANK			
	1	2	3	4
Low Pessimistic	D	C	A	B
Low Best	A	B	D	C
Low Optimistic	B	A	D	C
Low Average	B	A	D	C
Low Range	B	A	D	C

Figure 6

Although numerous statistical tests are available, they fail to provide conclusive results with available data at meaningful significant levels (.10, .05, or .01). Insufficient data is yielded by only "sampling" three times from a population. If a rigorous statistical test is desired the decision-maker could require eight individual Best estimates of risk for each alternative. It may not be illogical to assume that on any large program eight analysts can be found who could respond satisfactorily to the programmed questions. Eight estimates on each alternative would provide enough data to determine statistically which alternative offers minimum risk.

If it is assumed that the risk estimates were obtained from sampling normal populations, with unknown but equal variance, then the Students t-distribution can provide satisfactory answers for small samples by testing two alternatives at a time. It is recommended that the assumption of equal variance be further tested using standard F-distributions. If it is desired to test all alternatives simultaneously, and the previous assumptions are maintained, an Analysis of Variance can be performed.

Non-parametric statistics offers tests which can be just as powerful as those requiring the assumption of normality (7). For this reason, the Wilcoxon test for the unpaired case is recommended. In the Wilcoxon test samples from two populations which are tested against an hypothesis of equal means, are ranked numerically by size. If five estimates are obtained for each alternative then the sum of the ranks is

$$\text{Sum of Ranks} = \frac{10 \times 11}{2} = 55$$

A valid conclusion to be drawn from the null hypothesis is that the sum of ranks of each alternative must be approximately equal. Consider the following example using data from Figure 7.

ALTERNATIVE A		ALTERNATIVE B	
RISK ESTIMATE	RANK	RISK ESTIMATE	RANK
48	10	39	4
43	8	36	2
38	3	42	7
45	9	33	1
41	6	40	5
Sum of Ranks	36		19

Figure 7

There are 252 possible ways of selecting 5 ranks from a total o · ı ($C_{10,5}$=252). It is also possible to determine without muc. fficulty all feasible combinations summing to 19 or less. There . e 12 such combinations ranging from

$$
\begin{array}{ll}
 & 1 + 2 + 3 + 4 + 5 = 15 \\
 & \vdots \\
\text{to} & 1 + 3 + 4 + 5 + 6 = 19
\end{array}
$$

Accordingly, the probability, given the null hypothesis, of obtaining a sum of ranks $\leq$ 19 is

$$P = 12/252 = 0.0477$$

The null hypothesis could be rejected at the 5% significance level. Alternative B would be preferred.

While tables are available for calculating probabilities with larger sample sizes, the formal statistical approach, both parametric and non-parametric, is inherently limited by the difficulty of obtaining many qualified personnel to develop "best" estimates of risk. There cannot be much reassurance in relying upon this approach. The problem remains how to evaluate alternatives each with a single range estimate of risk.

There is intuitive appeal in the philosophy which proposes that quantitative results should only indicate the correct choice to the decision-maker - not cause him to doubt his strategy or change his goals. In other words, the decision-maker should have decided upon his criteria a-priori and not be swayed by how the numbers fall. That being the case, the conservative decision-maker may opt to pursue "the lowest of the Pessimistic" strategy - in essence, mini-max.

There exists persuasive arguments for selecting "the lowest of the Best" estimate. Certainly it should be considered at least as strong as either the Optimistic or Pessimistic. The decision-maker favoring the strategy of "lowest sum of estimates" would be confronted with the delemma, as seen in Figure 5, of choosing Alternative B over A even though the Best estimate for A is better than B. He could not be faulted for disregarding this fact and reaffirming his choice of B. However, he might resort to a sensitivity analysis to see how Alternative A would be more acceptable to him than B. For

this example a weight ratio of 3.0 to 1 in favor of the Best estimate over Optimistic or Pessimistic would represent the point where he is indifferent between both alternatives. This is less than the ratio of 4.0:1 proposed by Best (8) who favors the weighted average approach. In comparing alternatives A and B they would be measured accordingly:

A	B
75 x 1 = 75	80 x 1 = 80
40 x 4 = 160	45 x 4 = 180
30 x 1 = 30	10 x 1 = 10
265	270

Figure 8

and A selected as having the lowest weighted average. This procedure is based upon assuming a Beta distribution and is widely used throughout the Department of Defense.

In summary, the following recommendations may be made concerning evaluation or risk estimated:

1. Always check for dominance.

2. Whenever possible attempt to obtain sufficient (at least six) independent Best estimates of each alternative to enable a non-parametric statistical test to be accomplished.

3. If #2 is not possible, determine a strategy and make the appropriate decision. Attempt a sensitivity analysis to assess the limits to which the decision will remain unchanged.

REFERENCES

1. Dienemann, Paul F., Estimating Cost Uncertainty Using Monte Carlo Techniques, Rand Corporation, January 1966.

2. Major General John R. Guthrie, Speech to AMC Project Managers Conference, Washington, DC, February 1970.

3. Perry, Robert L., The Mythography of Military R and D, Rand Corporation, May 1966.

4. Marshall, A.W. and Meckling, W.H., Predictability of Costs, Time and Success of Development, Rand Corporation.

5. MacCrimmon, Kenneth R., Descriptive and Normative Implications of the Decision Theory Postulates, Western Management Science Institute, UCLA, April 1966.

6. Enthoven, Alain C., Systems Analysis and the Navy, Naval Review, 1965.

7. Alder, Henry L. and Roeseler, Edward B., Introduction to Probability and Statistics, W. H. Freeman and Company, 3d Edition, San Francisco, 1964.

8. Best, Guy A., Manageriel Direction and Control, Management: Concepts and Control, Industrial College of the Armed Forces, 1967.

VERT - A Tool to Assess Risk*

Gerald L. Moeller, P.E.
U.S. Army Management Engineering Training Agency
Rock Island, Illinois

VERT, an acronym for Venture Evaluation and Review Technique, is a mathematically oriented simulation networking technique. It is used to assist management in the decision-making process involving risk assessment of an on-going project or any new Government or business venture. VERT enables the user to create a fourth dimension of risk which is the common measure used to integrate the three principal dimensions of time, cost, and performance. With this technique, time, cost, and performance are the exogenous variables that control the values the endogenous variable "risk" assumes. An extensive array of operands facilitates the capability to model real time decision logic and enhances the exploration of conditional multivariate situations which defy ready mathematical analysis. Decisions within the network can be structured singularly or jointly on time, or cost, or performance basis. Classical discounting of cash flows formulas are also utilized.

VERT is a tool used for constructing cost, schedule, and performance analysis models of Government and business ventures. This tool is designed to systematically assess the risk involved in undertaking a new venture or in the planning, monitoring, and evaluation of on-going projects and programs.

During the last four to six years, business schools and a handful of companies have been teaching and using statistical decision tools to aid in project planning and review efforts. In the Defense industry, these efforts have been more formalized into a decision risk analysis function (Packard, 1970). The literature contains many examples of analysis and tools used in these efforts. Charnes (1966), Elmaghraby (1966), Pritsker (1966), Kaufmann & Desbayeille (1969), Schlaifer (1969), Raiffa (1970), Robinson (1970), Hwang (1970, 1971), Hwang & Banash (1971), etc. contain only a few of the many excellent examples of the current processes used by business and Government to reduce the uncertainty involved with a new venture or in monitoring an on-going project or program. VERT incorporates many ideas the above authors contributed for the development of a more formalized decision risk analysis.

I. Description of VERT Process

VERT is a network tool which utilizes simulation as a means of deriving solutions. It has an extensive array of logical and mathematical features which makes it possible to analyze complex systems and problems in a less inductive manner than traditional methods. When using this tool, the user can expend more time on individual component time, cost, and performance analysis rather than developing the interaction among components. The extensive number of operands available removes the inductive headaches from modeling component interaction. These operands enable the user to explore conditional nonlinear multivariate situations which defy ready mathematical analysis. VERT enables the user to create a fourth dimension, "risk," which is used as a common measure to integrate the three principal dimensions of time, cost, and performance. Risk is the endogenous variable being controlled by the exogenous variables time, cost, and performance.

VERT has two parts. Part one consists of constructing a graphic network representation of the project. Part two consists of analyzing that network through the use of a computer program.

Figure 3 is an example graphical network representation depicting elemental activities, events, and real time decisions. Real time in this context has the following connotation: the decisions made within this mathematical simulated network would be the same as those the manager on the job would make, given the time, cost, and performance values derived by the network for each of the various decision alternatives proved to be the same as those encountered in the actual project development.

In the VERT system, project activities are represented by arcs, and events or milestones are represented by nodes. The arcs and especially the

*VERT will be published in the 1972 Technical Papers of the American Institute of Industrial Engineers, Inc. and will be presented at the 23rd Annual Institute Conference and Convention of the American Institute of Industrial Engineers to be held in Anaheim, California, May 31 - June 3, 1972.

nodes are used to create the real time decision capability. Therefore, the flexibility and array of capabilities structured in the nodes and arcs become a very critical consideration when attempting to model an unusual decision situation. The author's intent is to strike a balance between having enough features available to efficiently model any decision situation versus over burdening the user with features to the point that only technicians can cope with this tool.

While pictorially describing the project in terms of the VERT operands, numerical values for time, cost, performance, achievement and event probabilities are assigned to the various project elements. Procedures useful for eliciting data have been suggested by Dalkey (1970), Northrop (1970), and Raiffa (1970). The numerical values assigned must be measured in a consistent manner throughout the network. Time cannot be laid out in terms of weeks in one section of the network and in terms of years elsewhere. Likewise, cost must be measured in identical units as ten, hundred, or thousand dollars, etc. throughout the network. Performance can be expressed in terms of any meaningful index such as horsepower, weight, reliability, utiles, return on investment, quality appraisal, systems worth, etc.

Time, cost, and performance for each activity can be jointly or singularly modeled as a functional relationship with other time, cost and performance parameters in the network and as a stochastic variable. This dual capability enables modeling the functional relationship portion of a regression equation among key parameters in the network and additionally modeling the stochastic residual. VERT has the following 14 transformations to aid in the task of expressing functional relationships among the key parameters.

No.	Transformation	No.	Transformation
1	$C_1X_1 \rightarrow X_2$	9	$C_1[\mathrm{Log}_{10}(C_2X_1)] \rightarrow X_2$
2	$C_1(1/X_1) \rightarrow X_2$	10	$C_1[\mathrm{Sin}(C_2X_1)] \rightarrow X_2$
3	$C_1(X_1 + C_2) \rightarrow X_2$	11	$C_1[\mathrm{Cos}(C_2X_1)] \rightarrow X_2$
4	$C_1(X_1 - C_2) \rightarrow X_2$	12	$C_1[\mathrm{Arctan}(C_2X_1)] \rightarrow X_2$
5	$C_1(X_1^{C_2}) \rightarrow X_2$	13	$X_1 \geq C_2$: $C_1X_1 \rightarrow X_2$ otherwise $C_1C_2 \rightarrow X_2$
6	$C_1(C_2^{X_1}) \rightarrow X_2$		
7	$C_1(e^{C_2X_1}) \rightarrow X_2$	14	$X_1 \geq C_2$: $C_1C_2 \rightarrow X_2$ otherwise $C_1X_1 \rightarrow X_2$
8	$C_1[\mathrm{Log}_e(C_2X_1)] \rightarrow X_2$		

X_1 represents a time, cost, or performance value previously derived within the network. C_1 and C_2 are inputted constants. C_1 is an ordinary multiplier of the transformed variable while C_2 is used to transform X_1 to X_2.

The functional modeling available in VERT will enable deriving time, cost, and performance values for each activity as a function of the following: (i.e., X_1 can be any of the following previously derived values) (1) node (event) time, cost, performance values (2) other arc (activity) time, cost, performance values (3) time, cost, performance of the given activity. (A parameter must not be dependent upon itself and there must be a dependency hierarchy established among these three principal parameters.) To aid stochastic modeling, VERT has 10 statistical distribution input options which are as follows: (1) constant, (2) uniform, (3) normal, (4) triangular, (5) erlang, (6) lognormal, (7) poisson, (8) gamma, (9) beta--3 or 4 parameters, or (10) any distribution, entered as a histogram approximation to the probability density function.

The degree or extent a project needs to be segmented into activities and events is a function of available data and the results desired. Some managers prefer to estimate parameters for entire modules or higher level work packages, rather than estimating parameters for the smaller elemental items in those work packages. Problem size sometimes has a bearing on the way the network is structured. If a problem is large, it is often advisable to construct lower level networks (subnets) of major modules. The histogram inputting capability for an activity's time, cost, and performance enables stochastic substitution of results from lower level subnetworks into a higher level network.

Part two of the VERT procedure consists of analyzing the network through the use of a computer program (Moeller, 1972). Networks are constructed so that various combinations of alternative activities could occur to make a project successful. The computer program explores alternate ways of completing the project through the technique of simulation. Upon simulating the network a sufficient number of times, the computer program prints out the following node time, cost (discounted, if desired), and performance information:

1. Pictorial histogram approximations to the probability density function.

2. Pictorial histogram approximations to the cumulative density function (see Fig. 1, cell data are printed on the page following the histogram printouts).

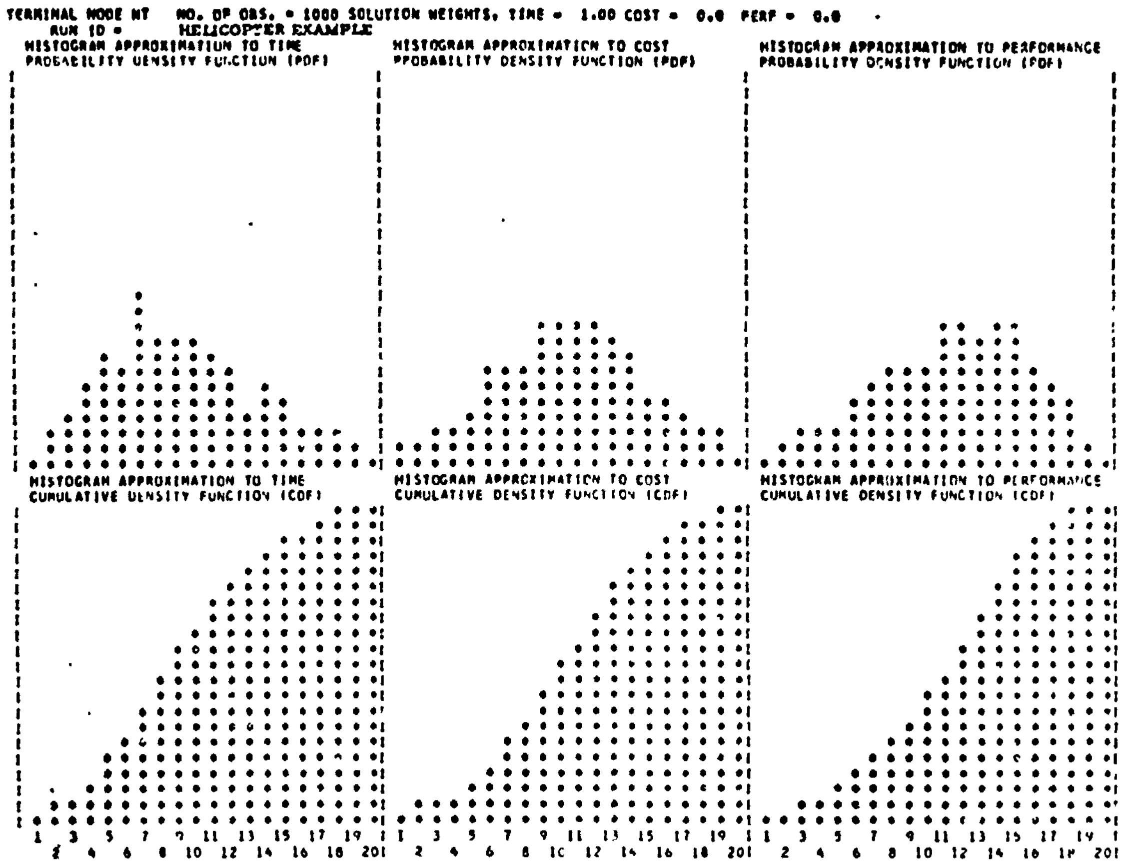


Figure 1. Histogram Printout

3. Mean observation.

4. Standard deviation.

5. Coefficient of variation.

This information is displayed for all internal nodes, intervals between nodes, and terminal nodes as requested. In addition, all terminal node time, cost, and performance data are combined to give a composite terminal node time, cost, and performance printout.

The histogram printout of the probability density function provides a picture of the range and concentration of time, cost, and performance values. Probability of exceeding certain value levels can be obtained from the histogram printout of the cumulative density function. The mean indicates the center of the distribution while the standard deviation gives an indication of the overall spread of the distribution. Lastly, the coefficient of variation enables an inference to be made on the spread of the distribution in relation to its mean.

VERT prints out a bar graph of terminal node utilization (similar to Fig. 2). It is through the use of this printout that the project "risk" can be ascertained. The usual form a decision risk analysis network takes is that of having one or several terminal nodes collect successful project completions, and one or several terminal nodes collect unsuccessful project completions. Realization of these various terminal nodes compared to the total number of iterations gives an indication of project success or failure. The program next prints out a critical path index for nodes (see Fig. 2) and arcs (similar to Fig. 2). Since different stochastic paths can be realized in the process of simulating the network, the critical path tends to change. Accordingly, the program computes the proportion of time each arc and node is on the critical path. These critical path options facilitate making sensitivity and crash program analysis.

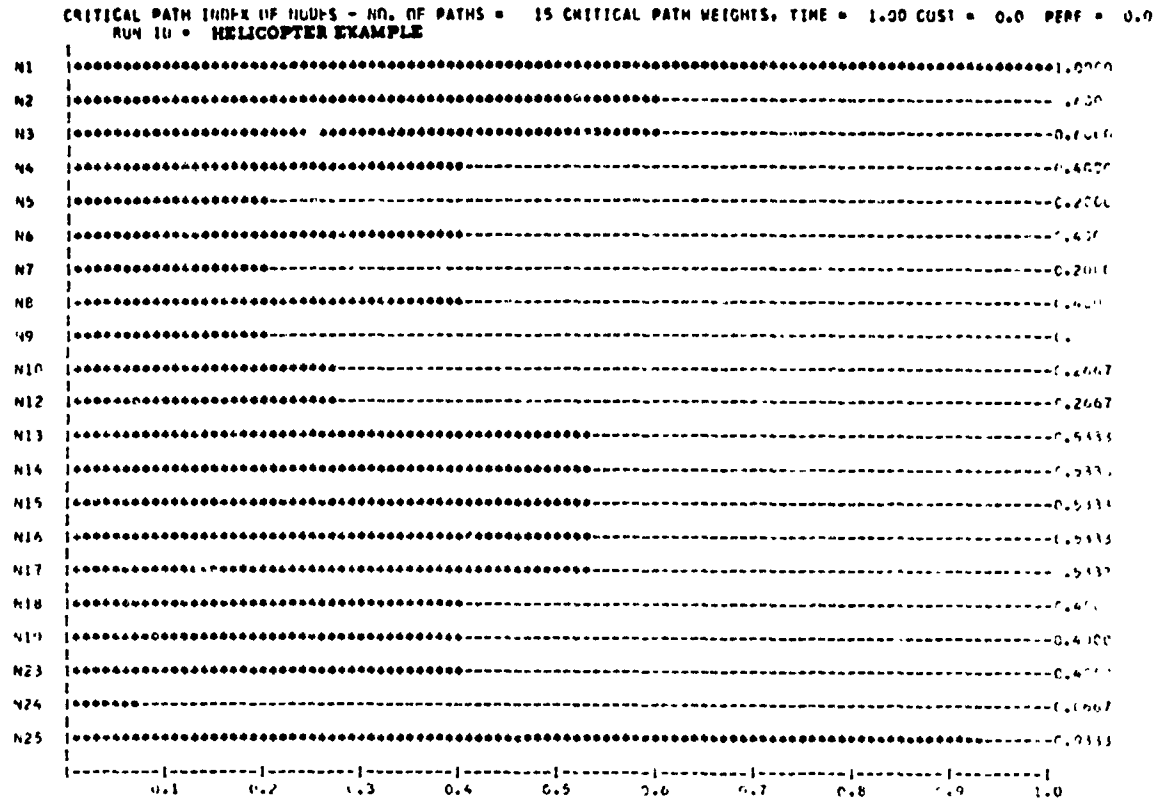


Figure 2. Bar Chart

II. Mechanics of the VERT Process

A. General Processing Steps

The processing steps of this program evolve around the various states the arcs assume in the processing sequence. These states are as follows: (1) uninitiated, (2) logically eliminated, (3) initiated, (4) successfully completed, (5) unsuccessfully completed, and (6) considered for the critical path.

The sequence of steps the program takes in deriving a solution for a single iteration is as follows: First, all arcs are given an uninitiated status. Initial nodes are processed next. They initialize or logically eliminate output arcs emanating from them. All initiated arcs are Monte Carlo processed to determine their success/failure status. Time, cost, and performance values are derived for each processed arc by the functional relationships and/or statistical distribution inputted for them. Next, all nodes are reviewed to determine if their input arcs have a processing status. Node processing includes deriving time, cost, and performance values via the input logic structure. The output logic for each partially processed node is employed to initialize and/or logically eliminate output arcs. This identical procedure is repeated again starting with all eligible arcs until the network flow has exhausted itself into the terminal nodes. At this point, all arcs should have been realized or logically eliminated from the network. The optimum terminal node is next determined as the one with the shortest completion time, lowest cost, or highest performance, or the best weighted combination of these three factors. VERT provides the capability to partially or fully cost the activities which were initiated before but not completed by the time the optimum terminal node was realized. If time, cost, and performance data displays for internal nodes were requested, the program now stores the necessary items to complete the displays. The critical path is next determined and stored as the path with the longest completion time, highest cost, lowest performance, or the least desirable weighted combination of these factors. VERT enables optional suppression of critical paths originating from certain

terminal nodes. The program continues to the next iteration repeating the preceding steps.

B. Operands

The basic building blocks (operands) of VERT are nodes and arcs. They are the vehicles used to express the unique aspects of a project. Their functional relationships are so interdependent that it is nearly impossible to describe the functions of one without describing some aspects of the other. Arcs perform two functions in the network: the primary function is to represent project activities and secondary to perform a logic function within the network. When an arc is used in this latter capacity only, it is referred to as a transportation arc. Every arc in the VERT system is characterized by the following:

1. An arc name
2. The name of its input node
3. The name of its output node
4. Probability of arc completion

Transportation arcs require only the preceding four attributes while arcs representing actual activities require some of the following items:

5. Separate equations (structured via the transformations built in VERT) for activity time, cost, and performance.

6. Stochastic variates for time, cost, and performance.

Nodes having Filter #1, #2, or #3, and time/cost/performance probability output logic, which are later discussed, require output arcs to carry the following additional information:

7. Filter #1 - upper and lower limits on time and/or cost and/or performance.

8. Filter #2 - upper and lower limits on the number of successfully completed input arcs.

9. Filter #3 - names of other arcs accompanied by an indicator.

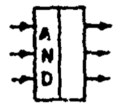
10. Time/cost/performance output logic-probability distribution(s) possibly requiring time/cost/performance boundaries.

There are four basic input logics available for the split-logic nodes. These logics are defined as follows:

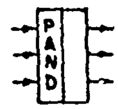
1. 

"Initial" - This input logic is used to start the network. Multiple initial nodes may be utilized in a network. Time, cost, and performance values assigned are zero.

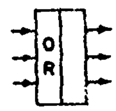
If the input logic for the following nodes is not satisfied, all output arcs will be logically eliminated.

2. 

"And" - This input logic requires all input arcs to be successfully completed before the network flow can continue through this node. The time value assigned to this node is the maximum path time of all the input arcs. Cost and performance values assigned to this node are computed as the sum of all the respective costs and performances of each input arc.

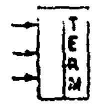
3. 

"Partial and" - This input logic requires all input arcs to be successfully completed or logically eliminated from the network. If at least one input arc has been successfully completed, network flow will be allowed to continue through this node. The time value assigned to this node is the maximum path time of all the successfully completed input arcs. Cost and performance values assigned to this node are computed as the sum of all the respective costs and performances of each of the successfully completed input arcs.

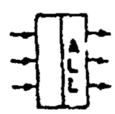
4. 

"Or" - This input logic requires all input arcs to be successfully completed or logically eliminated from the network. If at least one input arc has been successfully completed, network flow will be allowed to continue through this node. The time and performance values assigned to this node are the time and performance values carried by the input arc having the minimum path time. The sum of all the path costs of each of the successfully completed input arcs is the cost value assigned to the node.

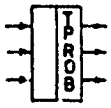
The following six output logics available for split nodes will be utilized only when the input logic can be successfully executed.

1. 

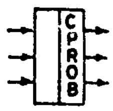
"Terminal" - This logic is used to end the network.

2. 

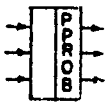
"All" - This logic will simultaneously initiate all output arcs emanating from this node.

3. 

"Time Probabilistic"

4. 

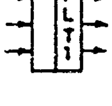
"Cost Probabilistic"

5. 

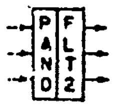
"Performance Probabilistic"

Each of the three preceding logics will initiate only one output arc. Arc initiation is accomplished probabilistically and can include a time/cost/performance basis if desired. The probability-time/cost/performance dependent situation enables inputting three different sets of output probabilities of initiation separated by two time/cost/performance boundaries. These boundaries create the three regions where the three probability sets apply. If the time/cost/performance computed for the node lies between zero and time/cost/performance boundary one, the appropriate time/cost/performance domain is region 1. Probability set number 1 will be utilized in this case. Likewise, if the node time/cost/performance lies between time/cost/performance boundaries 1 and 2, the appropriate time/cost/performance domain is region 2 and probability set number 2 will be utilized. Lastly, if the node time/cost/performance lies beyond the time/cost/performance boundary number 2, the appropriate time/cost/performance domain is region 3; probability set number 3 will be utilized.

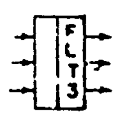
If time/cost/performance conditioning is not required, only probability set #1 needs to be specified. Likewise, if it is deemed that two probability sets separated by one time/cost/performance boundary fit the situation, a single time/cost/performance boundary point and probability sets #1 and #2 are required. The probability-time/cost/performance dependency capability is utilized in situations where the chances of events happening depend upon the time/cost/performance realized at key milestones within the network.

6. 

"Filter #1" - This logic will initiate one or a multiple number of output arcs. The criteria for initiating output arcs is based on jointly or singularly satisfying time, cost, and performance constraints placed on arcs emanating from this node. These constraints consist of upper and lower time and/or cost and/or performance boundaries. If the node time and/or cost and/or performance lies within the constraints placed on a given output arc, that arc will be initiated. Otherwise, the arc will be logically eliminated from the network. N-1 of the N output arcs must have constraints placed on them. The Nth output arc must be entirely free of any constraints. This arc functions as an escape arc in the event the constraints of the other output arcs have been violated. The escape arc will be logically eliminated from the network if at least one constrained output arc is initiated. Boundaries for the constrained output arcs can be (1) overlapping, (2) continuous, or (3) non-continuous, i.e., having gaps. This node can be processed with one, two, or three constraints simultaneously being employed. Most large-scale projects have time, cost, and performance constraints which should be observed. It is appropriate to use this logic to filter off those simulation iterations which do not fall within the limits of the time and/or cost and/or performance constraints.

7. 

"Filter #2" - This output logic is nearly the same as FLT 1 except for the following factors: (1) Only one constraint rather than one to three constraints can be placed on the output arcs. This constraint consists of an upper and lower bounds on the number of input arcs realized by this node; and (2) only PAND input logic may be employed with FLT 2 output logic. FLT 2 output logic is useful in constructing testing situations.

8. 

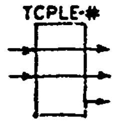
"Filter #3" - This logic will initiate one or a multiple number of output arcs. The criterion for initiating a given output arc is based on jointly satisfying all the arc completion constraints placed on it. Otherwise, the arc will be logically eliminated from the network. These constraints are specified by listing the arc names and attaching a plus or a minus to these names. If a plus is attached to the arc name, this constraining arc must have been successfully completed before the output arc being constrained can be initiated. If a minus is attached to the arc name, this constraining arc must have been unsuccessfully processed or logically eliminated from the network before the output arc being constrained can be initiated. N-1 of the N output arcs to this node must have from 1 up to 15 arc completion constraints placed on them. The Nth output arc is an escape arc which will be initiated only in the event the other N-1 arcs fail to be initiated. The escape arc will be logically eliminated from the network if one or more constrained output arcs are initiated. This output logic is especially useful for situations where successful completion of prior activities or

the failure of prior activities requires the initiation of other activities positioned farther on in the network.

For the preceding split logic nodes, time, cost, and performance values assigned to output arcs consists of the sum of the time, cost, and performance values derived for those activities plus the time, cost, and performance values assigned to the input node.

There are four special nodes having unit logic rather than having separate input and output logic. They require an indication of how many output arcs are desired to be initiated. This number is indicated in actual network drawings where the pound sign appears in the small pictorials accompanying these definitions.

1. **TCPLE-#**



"Time Cost Performance Link Escape" - This node has N input arcs coupling with one particular output arc. Additionally there must be one uncoupled output arc. This arc plays a role comparable to the role played by the escape arcs in the previously defined filter output logic. Execution of this node can be accomplished in one of two ways depending upon the input specification and the internal network action.

a. The number of output arcs initiated depends upon how many input arcs were successfully completed and on how many output arcs were desired to be initiated. One or all or a subset of all the linked output arcs may be initiated. If there are more successfully completed input arcs than there are output arc initialization requests, the following selection logic is utilized. Those output arcs will be initialized whose corresponding input arcs form an optimal subset. Optimal subset selection can be based on minimum total path time, cost, or maximum path performance, or the best weighted combination of these three factors. The remaining output arcs will be logically eliminated from the network.

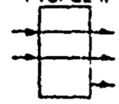
The time value assigned to this node is the maximum time value required by the most time consuming arc in the optimum input arc subset if time is used as the only decision criterion. If another decision criterion is used to select the optimum input subset, the node time value is recorded as the maximum time value of all input arcs which have not been logically eliminated from the network. Cost and performance values assigned to the node are computed as the sum of the cost and performance values of all input arcs successfully completed.

If the number of requests for output arc initializations cannot be fulfilled entirely, all output arcs will be logically eliminated from the network except the escape arc which will then be initiated.

b. If all the input arcs fail to be successfully completed, the escape output arc will be initiated.

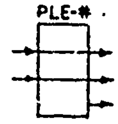
If the escape arc is initiated, the time value assigned to this node is the maximum time value of all input arcs which have not been logically eliminated from the network. Node cost is the sum of all path cost values of each of the input arcs which have not been logically eliminated from the network. The performance value assigned to the node under this failure condition is zero.

2. **PTCPLE-#**



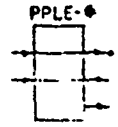
"Partial Time Cost Performance Link Escape" - This logic is the same as TCPLE except that the number of output arcs initialization request need not be fulfilled entirely. If one or more input arcs have been successfully completed, at least one corresponding output arc which links with a successfully completed input arc will be initiated. The escape arc will only be initiated when all input arcs to this node fail to be successfully completed.

3. **PLE-#**



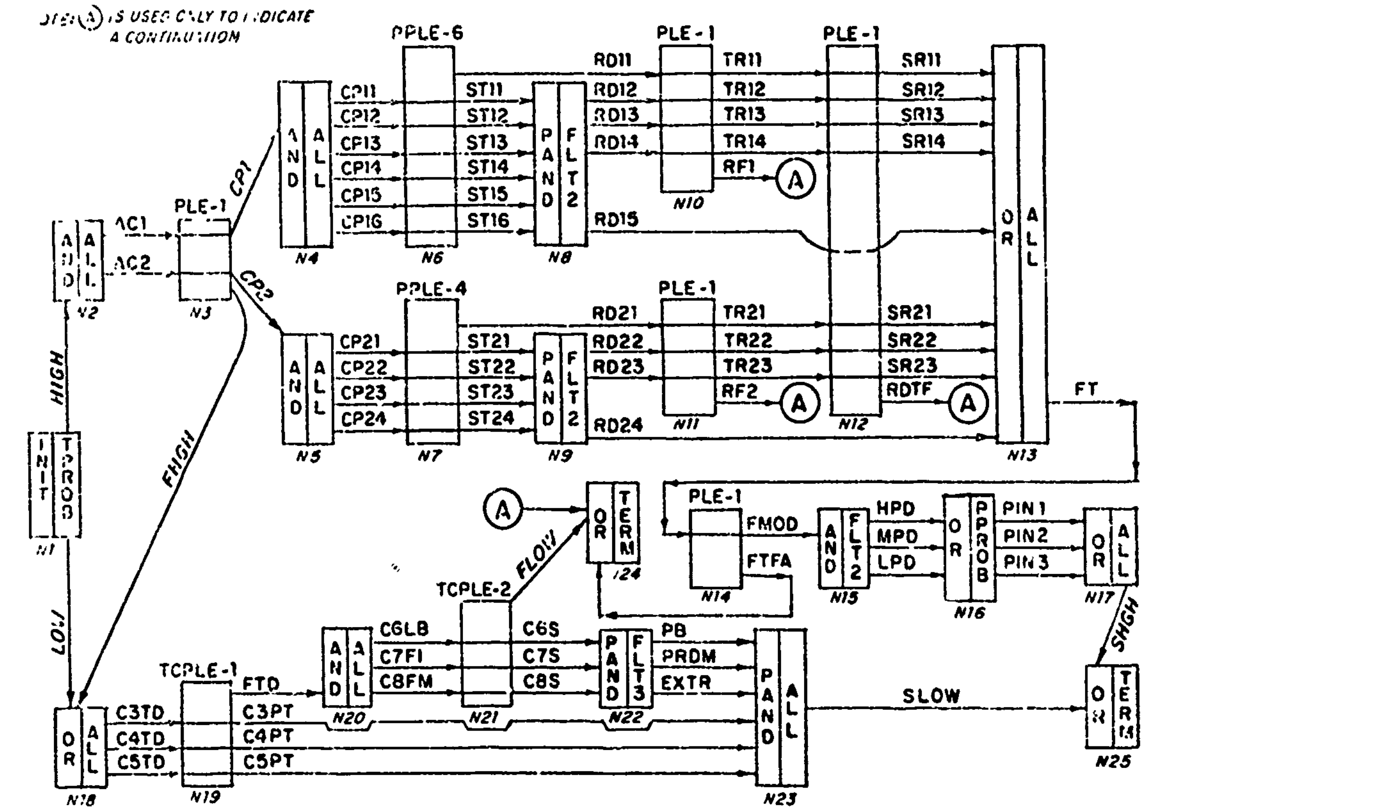
"Preferred Link Escape" - Has the same physical makeup as the TCPLE node. The only difference between these two nodes is the logic used to select an output arc for initialization. The logic in PLE requires that the first output arc be given preference over the second and the second be given preference over the third, etc. Thus, the criterion for selection is preference, not time, cost, or performance. Accordingly, the only thing that will prevent output arc number 1 from being initialized is that its corresponding input arc failed to be successfully completed.

4. **PPLE-#**



"Partial Preferred Link Escape" - This logic is analagous to PLE as PTCPLE is analagous to TCPLE. (The number of output arcs initialization request need not be fulfilled entirely.)

For the preceding special nodes, time, cost, and performance values assigned to output arcs are computed as the sum of the time, cost, and performance values derived for those arcs plus the time, cost, and performance values of the linked input arc. The escape arc is an exception to this rule. Its time and cost value is computed as the sum of the time and cost values derived for this arc plus the time and cost values assigned to the input node. Performance value for this arc is computed as the performance values derived for this arc.



HELICOPTER EXAMPLE *Figure 3*

Reproduced from best available copy.

These four nodes are especially useful for structuring major command or board of director type decisions. Since these nodes are the only ones which can accommodate input arcs having probabilities of completion less than one, they are often utilized as network flow continuity devices. In this capacity these nodes prevent the network flow from dying within the network.

III. Example Problem

The preceding mechanics and workability of VERT can best be illustrated through the network structuring of a hypothetical abbreviated example decision risk analysis problem as depicted in Figure 3. Suppose a large helicopter consumer is considering funding one of two possible helicopter development programs. Program #1 is a high-cost, large-scale funding program which consists of designing a complete new machine. This new machine will fill a gap in the present day array of machines available and, more importantly, serve as a steppingstone toward the development of a super ship deemed necessary ten years hence.

Two contractors have submitted proposals toward the development of this new helicopter. There are high risk problems in each of their preliminary designs which must be resolved through an Advanced Technical Components Study Effort (ATCSE) prior to follow-on design, prototype development, testing and production. Since contractor #1's design is preferred over contractor #2's design, contractor #1 will be awarded the follow-on effort if both contractors have a successful ATCSE. If contractor #1 does not have a successful ATCSE, however, and contractor #2 has a successful ATCSE, contractor #2 will be awarded the follow-on effort.

Both contractors' programs will be essentially the same after ATCSE, except the design offered by contractor #1 is considered to be of greater risk. Therefore, contractor #1 will be required to build and initially test six prototypes, while contractor #2 will be required to build and test four prototypes. If none of the prototypes fail, the design will be considered a success and some modifica-

tions will be made to correct weak areas elicited in the initial test phase. In the event one or more prototypes fail, various degrees of redesign effort would follow depending upon the number of prototype failures. Additionally, each of these redesign efforts has risk associated which could cause project failure. Following redesign, some in-house testing will be conducted which may produce conditions that could again cause project failure.

At this stage, the contractor should have a prototype ready for the most challenging test of all--actual on the job field testing. This phase of the program could also produce a project failure. However, if the field testing is successful, some final touch-up modifications will be made prior to mass production.

The number of helicopters built will be a function of the time elapsed and the development cost incurred prior to production. Currently, three levels of production are under consideration. Prior studies have indicated that repair parts inventory is a function of the performance achieved in the overall design, development, and testing of the machine prior to production. The higher the level of machine performance, the less inventory needed. Three levels of inventory treated in a stochastic manner would adequately model the phenomenon involved.

Program #2, in this hypothetical example, is a low-cost, small-scale funding effort. It will be started if insufficient funds are provided for the high cost program or if both ATCSE fail. This program consists of modifying the equipment currently being used in the field which meets the minimal requirements except the lifting requirement. This requirement can be met by increasing the motor's power or by developing a tandem motor arrangement. Engineers prefer trying the tandem approach before resorting to the power development. Since the tandem design is considerably riskier, three tandem design efforts will be funded. The contractor with the least cost design will be awarded follow-on production.

If the tandem effort fails, three private consultants will concurrently be retained to attempt to devise a means of increasing the power of the present motor. One of these consultants will work on a large bore design, another on fuel injection techniques, and the third will work on fuel mixtures. Each of these areas of study has some chance of failure. The coupling of any two of these three developments should yield a sufficient power improvement to make a successful machine. To speed the process, the first two consultants who successfully complete their developments will be jointly awarded a bonus. The present manufacturer of the helicopter will use these revisions in a new major production effort. In addition, if the fuel mixtures method is utilized, a small field fuel blender will have to be manufactured.

The graphical pictorial network (Fig. 3) for this problem consists of two major parts: the top part depicts program #1--the high cost option (nodes N2-N17); the bottom part depicts the low cost program (nodes N18-N23). Node N1 starts the network flow by randomly directing a certain proportion of the iterations as inputted toward the high cost option. For the remaining iterations, the flow is directed toward the low cost option. Arcs HIGH and LOW are example transportation arcs. They do not represent actual activities, rather they function as network flow transportation devices.

Node N2 is used to initiate arcs AC1 and AC2. These arcs represent the ATCSE of contractor #1 and #2 respectively. Since these arcs represent risky activities, they have probabilities of successful completion which are less than 1. AC1 and AC2 flow into node N3. This node has PLE logic requiring only one linked output arc to be initiated for normal completion of the node. Node N3 determines which of its output arcs will be initiated on the basis of the success of these two ATCSE and the preference logic structured in this node. If arc AC1 is successfully completed (successful/unsuccessful status determined by the Monte Carlo process described in Section II A), arc CP1 will be initiated regardless of the successful/unsuccessful status of arc AC2. The remaining output arcs CP2 and FHGH will be given a status of logical elimination from the network. If AC1 fails and AC2 is successful, arc CP2 will be initiated and the other output arcs will be logically eliminated. If AC1 and AC2 both fail, the high cost option for this helicopter development effort will be abandoned. Transport arc FHGH will be initiated and the remaining output arcs will be logically eliminated.

Arc CP1 represents contractor #1's design and construction of six prototypes. Likewise, arc CP2 represents contractor #2's design and construction of four prototypes. In an actual project, arcs CP1 and CP2 and similar arcs could represent the results of extensive subnetworks of their own. Results obtained from independently processing these subnetworks can then be stochastically substituted back into this network in the form of arcs CP1 and CP2.

Arcs CP11, CP12, CP13, CP14, CP15, and CP16 represent the risky activities of testing the six prototypes constructed by contractor #1. These activities are funneled into node N6 which has PPLE logic requiring at least one and allowing up to six of the linked output arcs to be initialized. The remaining uninitiated arcs will be

logically eliminated from the network. Total failure of all these input arcs results in the initiation of the escape arc RD11 and the logical elimination of the remaining output arcs. Output arcs ST11, ST12, ST13, ST14, ST15, and ST16 are transport arcs which feed into node N8. This node is a selection device used to determine what level of redesign effort should follow testing. The level selected by the FLT 2 output logic is a function of the number of successful input arcs realized by this node.

Arcs RD11, RD12, RD13, RD14, and RD15 represent a spectrum of redesign activities. RD11, at the low end of the spectrum, is a high cost activity having a low probability of successful completion. This arc is realized only when all the prototypes tested fail. RD12 is a moderate version of RD11 but still costly and risky. It will be initiated when either four or five of the prototypes fail. RD13 is initiated when two or three prototypes fail and RD14 is initiated when one prototype fails. Lastly, RD15 is initiated when none of the prototypes tested fail. This arc represents a minor modification activity not having any chance of failure. Consequently, arc RD15 travels directly into node N13. The remaining redesign arcs (RD11, RD12, RD13, and RD14) all flow into node N10 which has PLE logic requiring one linked output arc to be initiated for normal processing.

Output arcs TR11, TR12, TR13, and TR14 represent testing of improvements made during the corresponding linked redesign effort. If all the input redesign efforts to node N10 fail, transport arc RF1 will be initiated. RF1 flows into failure node N24. Second round test activities TR11, TR12, TR13, and TR14 all have probabilities of completion smaller than one. These arcs flow into N12 which has PLE logic requiring one linked output arc to be initiated for normal processing. Output arcs SR11, SR12, SR13, and SR14 are transport arcs traveling directly into node N13. This node acts as a vertex for contractor #1 and #2's programs, since their programs are identical for the balance of the high cost option.

The portion of contractor #2's effort covered by the activity between nodes N5, N7, N9, N11, N12, and N13 is similar to contractor #1's efforts covered by the activities between nodes N4, N6, N8, N10, N12, and N13. Because of this similarity the network estimate for preceding mentioned efforts of contractor #2 will be left to the reader. It should additionally be noted that PTCPLE logic could have been used in place of PPLE logic for nodes N6 and N7. TCPLE, PTCPLE, and PPLE logic could have been used in place of PLE logic for nodes N10, N11, and N12.

Arc FT represents field testing of the reworked prototypes developed by either contractor #1 or #2. Node N14 receives the network flow from arc FT and routes this flow to arc FMOD or to arc FTFA depending upon whether FT is a success or a failure. Node N15's FLT 1 output logic is structured to select a level of production (represented by output arcs HPD-high production, MPD-medium production or LPD-low production) as a function of time expended and cost accumulated up to this node. Node N16 randomly chooses a level of field parts inventory (represented by output arcs PIN1, PIN2, or PIN3) according to an ad hoc distribution developed from prior parts inventory studies. However, the distribution that applies would change according to the amount of performance generated in the network prior to the realization of this node. The more performance generated, the less need for a large parts inventory. Three distributions separated by two performance boundaries will need to be inputted. The performance boundaries mark the shift from the region where one distribution applies to the region where another distribution applies. Arcs PIN1, PIN2, and PIN3 represent high, medium, and low levels of parts inventory, respectively. The distribution which applies to the region up to the first boundary should be weighted toward realizing arc PIN1. The distribution which applies to the region between the first and second performance boundaries should be weighted toward realizing arc PIN2. Lastly, the distribution which applies to the region after the second performance boundary should be weighted toward realizing arc PIN3.

Node N17 will be realized after either arcs PIN1, PIN2, or PIN3 have been successfully completed. Transport arc SHGH will be initialized and processed after N17. This arc carries the network flow down to the success node N25.

Program #2, the low cost program, is initiated with the realization of transport arcs LOW or FHGH. These arcs flow into node N18 which initializes the three competing tandem motor design efforts represented by arcs C3TD, C4TD, and C5TD. The contractor who completes the least cost tandem design will be awarded follow-on production represented by arcs C3PT, C4PT, or C5PT. If all three contractors fail, transport arc FTD will be initiated. Node N19 functions in two capacities. First, it represents the action of an evaluation team which will study and cost out each of the contractor's prototypes to determine which one has the least cost design. Second, it will function as a switching device in the event all the tandem design contractors fail to make a successful prototype.

Node N30 initiates arcs C6LB (large bore design effort), C7FI (fuel injection effort), and C8FM (fuel mixtures effort). Node N21 rewards the two

consultants who successfully complete their effort first with a bonus represented by arcs C6S, C7S, and C8S. If two or more consultants fail, transport arc FLOW will be initiated. This arc carries the network flow to failure node N24. The production of a modified helicopter is represented by arc PRDM. This arc's initiation is contingent on the logical elimination of arc FLOW. Arc PB represents the production of a field fuel blender. This arc's initiation is contingent on the successful completion of arc C8S. Arc EXTR will never be initialized and is included to satisfy FLT 3 output logic of Node N22. If node N23 is realized, it will initiate transport arc SLOW. This arc carries the network flow into the network success node N25.

While constructing the logic just described, the user should also be accumulating input data. This hypothetical example may require the use of all the statistical distributions embedded in this program. It might also require the use of functional relationships. For example, the time spent on secondary test efforts TR11-TR14 and TR21-TR23 might be inversely related to the time spent on the primary testing efforts CP11-CP16 and CP21-CP24. Cost incurred in the design of the tandem motor might have an inverse effect on the manufacturing time. These are just a few of the many relationships which might exist in an actual development.

IV. Conclusion

A thorough validation of network logic, output results obtained, etc. should be made after computer debugging the network. Key personnel should check the authenticity of the entire project to date and should suggest the many trade-off considerations required of a thorough "decision risk analysis." The flexibility of the VERT operands enables the user to readily make the network modification which trade-off analysis requires. Exercising the technique enables the identification of the following:

1. Potential problems
2. Consequences of element failure
3. Low risk program areas
4. High risk program areas
5. Requirements versus state-of-the-art trade-offs
6. Adequacy of acquisition time
7. Sufficiency of appropriations
8. Insufficient performance
9. Optimum allocation of funds
10. Optimum allocation of time
11. Optimum allocation of design effort
12. Data gaps/recommend studies and concepts
13. Sensitive/critical parameters

Additionally, use of a formalized tool like VERT tends to generate particular interest among all operating personnel and engineers. This tends to heighten the level of participation and make people give more thought to subjective inputs. Thus, the quality of input becomes much higher.

Risk analysis is by nature an iterative process (USAF etc., 1971) and must be updated and validated at regular intervals. It should be coordinated with key decision points of the program life cycle. Also a timely risk analysis can be used as a basis for budget appropriation purposes. In summary, the basic jective of risk analysis study is to create a quantitative experimental laboratory to study program success. This laboratory then becomes a baseline toward controlling the ever present program problems of cost growth, schedule slippage, and degradation of performance.

Bibliography

Charnes, A., et al., "DEMON: Decision Mapping via Optimum Go-No Networks - A Model for Marketing New Products," Management Science, July 1966.

Dalkey, N., Brown, B., & Cochran, S., "The Delphi Method IV: Effect of Percentile Feedback and Feed-in of Relevant Facts," Memorandum RM-6118-PR, Rand Corporation, Santa Monica, California, March 1970.

Elmaghraby, S. E., "On Generalized Activity Network," Journal of Industrial Engineering, November 1966.

Hwang, J. D., "Analysis of Risk for the Materiel Acquisition Process-Part I: Fundamentals," Technical Report SY-R6-70, Systems Analysis Directorate, Hq., U.S. Army Weapons Command, Rock Island, Ill., November 1970.

Hwang, J. D.. "Analysis of Risk for the Materiel Acquisition Process-Part II: Utility Theory, Technical Report SY-R2-71, Systems Analysis Directorate, Hq., U.S. Army Weapons Command, Rock Island, Ill., May 1971.

Hwang, J. D. & Banash, R. L., "An Introduction to Design Analysis," Technical Report SY-TN1-71, Systems Analysis Directorate, Hq., U.S. Army Weapons Command, Rock Island, Ill., Feb. 71.

Kaufmann, A., & Desbayeille, The Critical Path Method (Gorden and Breach, 1969).

Moeller, G. M., "VERT," Documented in Army Automated Data Systems Inventory RCS CSOCS-(OT) 152 AMCMS Control No. 'G 55-1-204, Program #T008N1202. U.S. Army Management Engineering Training Agency, Rock Island, Ill., 1972.

Northrop, G.M., Davis, E.L., Sweeten, E.R., & Merlin, Cdr. W.F., "Use of Quantified Expert Judgement in Cost Effectiveness Studies," Presentation at the 38th National Meeting of the Operational Research Society of America, 1970.

Packard, David, "Weapon System Acquisition Policy, Army Research and Development News, July-August 1970.

Pritsker, A.A.B. & Happ, W.W., "GERT: Graphical Evaluation and Review Technique-Part I Fundamentals," Journal of Industrial Engineering, May 1966.

Raiffa, H., DECISION ANALYSIS Introductory Lectures on Choices under Uncertainty (Reading, Mass., Addison-Wesley, 1970).

Robinson, S.L., Preliminary Edition, "MATHNET -MATHEMATICA's Network Analyzer Program Princeton, N.J.: MATHEMATICA, November 1, 1970. Material included in a course presented by the Army Logistics Management Center (ALMC) titled "Risk Analysis."

Schlaifer, R., Analysis of Decision Under Uncertainty (New York, McGraw-Hill, 1969).

USAF Academy Risk Analysis Study Team, "Final Report of USAF Academy Risk Analysis Study Team," sponsored by Deputy for Systems, Aeronautical Systems Division (AFSC), Wright-Patterson Air Force Base, Ohio, 1 August 1971.

Reproduced from best available copy.

ENVIRONMENTAL RESEARCH IN COLD REGIONS

by

ANDREW ASSUR
U. S. Army Cold Regions Research and Engineering Laboratory
Hanover, New Hampshire

The history of warfare has shown that in addition to military strategy there are these factors which determine failure or success: Technology, the Environment and Psychological Factors. Opinions vary which of these four are the most important; there certainly have been cases of brilliant strategy, the United States in particular relies heavily upon superiority in technology but there are also creeping doubts whether sufficient attention has been paid to the remaining two factors.

The Corps of Engineers recognized the need for environment oriented engineering earlier than other services and initiated a multidiscipline oriented research program in the most difficult of the earth environments: the cold regions. A similar program in deserts and tropical areas has never materialized although it was almost started. Including the winter regime of temperate regions the program is relevant to vast land masses of the Northern Hemisphere, includes a good portion of the United States, Alaska, Canada, virtually the entire Soviet Union and a portion of Europe, a sizable part of China, the northern part of Japan and, of course, the Arctic Ocean and Greenland. It is also relevant to high mountains, such as the Himalayas, the Rocky Mountains, even isolated areas in Africa, Australia and South America. The Antarctic is the dominant feature in the South.

The research and engineering program of the U. S. Army in Cold Regions is unique in the Western World. It is still exceeded several times by Soviet efforts and it is supplemented in the Free World by a similar Institute in Japan. Even China is known to make sizable efforts. Many other civilized countries make at least a modest effort, Canada more than others.

The American program ranges from basic research into the properties of frozen materials and fundamental engineering research in soils, to research in excavations, foundations, pavements, structures and utilities and research supporting the military engineer in field activities. A sizable benefit is also being derived for the Civil Works activities of the Corps and for numerous industrial applications which especially at present are becoming increasingly important. Operational experience in cold regions leads also to advances in technology. The industrial breakthroughs in Northern Regions which are now on the verge of materializing are in part due to the forward vision of Army Engineers in coping with existing or anticipated difficulties.

The program has led to the ability to provide effective support for large scale operations of the Air Force and Navy in Cold Regions.

As a result the U. S. Army finds itself in a position to achieve better preparedness than any military service in the past, although a tremendous amount of work still must be done to achieve complete reliability and confidence to operate anywhere in the World including the Cold Regions.

FORECASTING SOIL TRAFFICABILITY FOR TUNISIA IN 1943

by

FINN E. BRONNER

U. S. Army Research Office Durham
Durham, North Carolina

During the Tunisian Campaign, HQ II Corps requested from AFHQ periodic forecasts of the "going" for the combat area. The terrain element of AFHQ Intelligence Division prepared an arbitrary map of the northern half of Tunisia, in which soils were grouped on the basis of several assumed moisture-retention properties and conditions, and the units keyed into a simple behavioral classification. Using the weekly ten-day weather forecast for Tunisia, prepared for use in operational planning on a priority basis, by the Army Air Corps Weather Detachment stationed in Algiers, the terrain element interpreted the probable condition of each soil unit for each day of the forecast period. The finished trafficability forecast was then transmitted as a priority message to HQ II Corps in Tunisia. Proof that this system worked, came to AFHQ in the form of a number of enthusiastic expressions of approval from division and regimental level.

ATMOSPHERIC EFFECTS ON ARMY OPERATIONS

by

MARVIN DIAMOND

U. S. Army Electronics Command
White Sands Missile Range, New Mexico

Army operations are significantly, often critically, affected by the physical presences of the atmosphere. Air operations, ground trafficability, artillery fire, chemical operations, target acquisition, radio communications, and personnel detection are all significantly affected by the conditions of the atmosphere during the time of passage of the projectile, vehicle, agent, signal, or effluent through the atmosphere. Almost without exception, the state of the atmosphere must be known before the event in order to permit adaptation to the environmental conditions which will affect Army operations. A continuous real time surveillance of the battlefield atmosphere is required to permit efficient operation of weapon, detection, and target acquisition systems. Atmospheric effects on these systems are discussed, and means for providing real time observations of atmospheric parameters are briefly described.

Preceding page blank

TERRAIN CONSIDERATIONS AND OPERATIONS RESEARCH MODELS

by

WARREN GRABAU
U. S. Army Waterways Experiment Station
Vicksburg, Mississippi

One of the major reasons for the reluctance to incorporate realistic terrain scenarios in operations analysis models is the difficulty of providing the requisite terrain information. It is quite true that this has been a formidable problem. And, since the existence of that problem has been one of the driving forces behind the work of the Waterways Experiment Station, it seems appropriate to take this opportunity to expand somewhat on the opening remarks.

Historically, there have been two hangups in the matter of terrain description in a form suitable for mathematical manipulation; one philosophical and one technological.

The philosophical problem stems from the fact that mathematical manipulation obviously requires that all elements of the problem be stated in numerical or quantitative terms. For such an obvious truism, that statement has caused an almost incredible amount of trouble. The reason is that the mathematicians went to the "experts" on terrain, namely the geographers, geologists, and agronomists. And the geographers, geologists, and agronomists didn't know what the mathematicians were talking about, because their disciplines had evolved on the basis of qualitative description. To be sure, there were statistics, which they all used, but it turns out that terrain description by standard statistical methods is very intractable indeed; there is too much local variability, and there is too much difficulty in identifying the populations which one is trying to describe statistically, and so on. The inevitable result was that very little of the statistical descriptions was of any use to the operations analysis types. And so it was concluded that the problem was not solvable.

The technological problem emerged from the fact that nearly all military activities are affected by a large number of terrain attributes or factors. It was quickly found, when one viewed the situation activity by activity, that the terrain descriptions required for even relatively simple tactical or strategic problems became almost incredibly complex. People took one look at the complexity and concluded that there was no hope of getting the necessary data in the first place, and if by some magic one could get it, that it would require so much computer storage that it could not be handled. So again the problem looked as if there was no solution.

But times change. There is now a new generation of terrain analysts who have grown up with mathematics, and it turns out that there are ways other than purely statistical to describe terrain. One can, for example,

Preceding page blank

describe a topographic surface deterministically, and derive from that description any statistical parameter required by an operations analysis model. One can do the same with vegetation, and so on. Thus, it looks as if the philosophical problem is being avoided by time and new education processes, and the technological problem is being side-stepped by a new way of looking at things.

These new ways are exemplified by things such as the so-called "factor maps", by computer graphics, by deterministic terrain-vs-activity interaction models, and so on. All of this leads to the conclusion that operations analysis models are at least potentially capable of becoming themselves deterministic in the way in which they simulate the effects of terrain on the subject being modelled.

MINIMIZING ENVIRONMFIITAL RISKS WITH WELL DESIGNED TOPOGRAPHIC PRODUCTS

by

ALEXANDER R. PEARSON

U. S. Army Engineer Topographic Laboratories
Fort Belvoir, Virginia

The vital link between the intelligence community and operational elements is the medium or product used to communicate the intelligence. Unless these products - be they paper maps, computer printouts, electronic displays, or voice messages - present needed information in a form that beneficially influences a user's decisions, they achieve little and the labors of the researchers come to naught. Only recently have the military services recognized the importance of good product design in strengthening this fragile link between users and holders of environmental data. USAETL has now established a Topographic Products Design Branch and undertaken a research and development program in the area. During past development efforts, the Laboratories have evolved concepts and practical approaches which will be applied to the development of a wide variety of general and special purpose products. Stated quite simply, the approach will be to determine the information content of a product and then design a presentation.

The underlaying design concepts proceed from the assumption that intelligence itself does not directly influence operations; it operates indirectly through actions of the users. Products must, therefore, be designed to influence user decisions. They must influence decisions made during the performance of identifiable tasks. Operations analysis will identify the tasks supported and the discrete elements of environmental data required. Psychophysical investigations will provide insight into the manner in which the users perception of the works, symbols, and images of the products affect his decisions and task performance.

Analysis on the Infrared Night Observation Device, Dr. Edward F. Allard, U.S. Army Electronics Command, Night Vision Laboratory

A. Introduction

1. Background

An existing two phase NODLR-IR contract was at a decision point at the Night Vision Laboratory. The decision was whether or not the Night Vision Laboratory should take its option to continue Phase II with Glorious Systems Inc. It appeared at that time that the NODLR system of Phase II design would be of unacceptable costs to the Army. Since technology in the infrared field was (and still is) advancing rapidly, it was uncertain if a new low cost NODLR could be developed at that time. A group of professional scientists and engineers with some operations research background was given the job of assisting the management in forming a set of decision alternatives for NODLR.

2. A Philosophy

There were four general objectives of setting up an inhouse systems analysis program for NODLR:

a. Supply management with sufficient information for decision making.

b. Apply the sound principles of systems analysis to NODLR.

c. Develop a new method of data gathering with verification.

d. Encourage management to apply the systematic approach to all programs.

To understand the minds of the men performing the analysis, these general objectives will be discussed briefly:

a. Supply management with sufficient information for decision making.

The problems we had to answer were: "What are the risks associated with developing a NODLR, which meets the Army's requirements, is of acceptable costs, and will be in the field in a reasonable time?" In comparison to large programs, such as building a destroyer or an antiballistic missile system, our problem was relatively simple. We were aware that there exists a 1001 ways to diagonalize a matrix and an infinitude of curve fitting programs. Also, operation research reports contain many dazzling presentations of results, but unfortunately most of these results leave the manager with stars in his eyes. In many cases there are pages of results with little information being passed on to the manager. We decided to keep the analysis as mathematically simple as possible, hoping that we could maximize the information flow to the decision makers.

Preceding page blank

b. Apply the sound principles of systems analysis to NODLR.

All the men involved in the study had used the scientific approach as part of their life. However, some people not in the scientific field do not use the systematic approach to problem solving in a formal way. Even though we felt that the whole subject of risk analysis was an exercise in common sense, we felt it was necessary to define our terms and formally establish a procedure. The procedure we followed was:

1. Define the NODLR problem. In this case, defining the problem was easy. We had contract information on an existing NODLR system.

2. Clearly state the objectives of the analysis with the boundary conditions or limitations. This was no trivial step since no analysis of this type had ever been done at NVL. It was not a matter of simply copying or improving some other report, but of starting from the beginning.

3. Configure the final report. This step assured us that our data gathering procedure would provide us with the necessary information for the final report.

4. Write a detailed plan of how the objectives would be met. The procedure followed in this step was completely new as far as we were concerned. Details will be fully explained later.

5. Perform an evaluation during the analysis to determine if the objectives were being met. It could happen, and did happen in fact, that the objectives listed in step 2 and executed in step 4 would not be met. This necessitated a change or modification in objectives with an associated change in plan. Borrowing the nomenclature from electrical engineering, this was essentially a feedback loop. Since the approach had never been tried, we were not sure that answers could be obtained.

6. Re-evaluate our approach based on the feedback of step 5. In some cases an objective had to be dropped and in other cases modified.

7. Draw conclusions.

c. Develop a new method of data gathering with verification.

There is no doubt that the track record for predicting development and production costs for Department of the Army systems is very poor. New approaches were devised to improve the track record. As far as we were concerned, the solutions were essentially the discovery of a new way to diagonalize a matrix. The problem was not being solved but rather neutralized. We saw the problem in all its horror - BAD DATA. We felt that an orangutan could turn the computer crank once he had GOOD DATA. So, we spent 75% of our time

gathering accurate input data and establishing a data bank for infrared systems. The method was new, time consuming, physically tiring and nerve wracking.

d. Encourage management to apply the systematic approach to all programs.

Since the average manager is hard working with limited time for new managerial techniques, we decided to write the final report in a heuristic fashion with all terms and techniques defined. We decided to spare them from mental acrobatics and drive dramatically to the point. We felt a complicated report would be the death knell of meaningful risk analysis.

B. Risk Analysis of the Thermal Night Observation Device Long Range (NODLR)

1. Introduction

The problem was to determine the risks associated with developing a NODLR, which would meet the Army's requirements, be of acceptable costs, and be in the field in a reasonable time. The most important boundary condition was that the Army would not accept an "expensive" systems. A report from the congressional record, predicting a production cost of $35K, was known to us. Since there were no violent objectives to this prediction, we used it as a ball park figure. By the way, its the first question every company asked us.

2. Objectives of the analysis

The specific objectives of the analysis were:

a. To determine if Phase II of the present NODLR contract should be continued.

b. To determine whether or not a new NODLR could be developed which would meet the Army's requirements, be of acceptable cost and not seriously delay getting the equipment in the hands of the troops.

c. To reduce the change of being over optimistic in the development and production cost of a new NODLR.

d. To quantify the problem areas in time, cost and performance with their associated risks being identified and quantified.

e. To show the desirability aspects of a truly open procurement process.

f. To objectively evaluate the contractors involved in the study.

g. To reduce data to an understandable form and present to management in an unbiased form.

h. To write a report which clearly explains the procedure of risk analysis.

i. To outline a method of control of the program after award of the contract.

Each of these objectives were met in the analysis. The most important aspects of the analysis will be given in this paper. A complete report on the study is in the process of being published.

3. Plan of the Analysis

a. Standardization

In any scientific set of experiments a standard of some kind must be established to compare data. Since this study was an attempt to use scientific methods in management, a standard for data collection that was realistic and yet not too complicated had to be established. If too many inputs must be considered, then cause and effect relationships may not be traceable. So, to prevent the analysis from turning into a Mulligan's stew certain limitations on inputs into the program were imposed. The effects on time-cost relationships due to types of contracts, efficiencies of companies, labor rates in different areas, and financial conditions of companies, are difficult to establish in a quantitative way. A decision was made to go to the companies themselves to gather information. A realistic model of a NODLR system was broken into components. It was felt that finer detail other than these components was neither necessary nor desirable. Since a comparison of companies would be necessary, the standardizing vehicle was to be a detailed questionnaire. Two methods of administering it were considered. One was to send the questionnaire to the companies, the other was to administer it in a face-to-face situation. It was decided to use the face-to-face method and at the same time tour the facilities and talk with various technical people. The data was documented carefully with names and places. In addition to gathering data for the NODLR program, this procedure provided a means of building a data bank for future analyses. Confidence in the data was enhanced by on-the-spot inspections.

Since this type of risk analysis had not been performed previously by the Night Vision Laboratory and was new to the infrared companies, it was important for standardization that the same government representative introduce the program to all concerned companies. All the details of the program and reasons for introducing this type of analysis were explained fully to all concerned with this study.

b. Errors

There are two kinds of error possible in an analysis of this type, random error and systematic error. Random errors refer to those errors that are caused by mistakes in evaluating a particular thing. For example, if all companies charge $10,000 for FIR optics and one company charges $25,000, then there is a good change that the $25,000 figure is in error. This type of error was minimized in our study. Systematic errors refer to those errors that are caused by errors of omission and constant type errors. For example, all companies might charge $10,000 for FIR optics. But not included on this figure is a $2,000 handling charge. This point might be overlooked causing a systematic error. This type of error is not important when comparing companies, but is important when absolute values are needed. This type of error is difficult to eliminate. There is a possibility that at least one systematic error is contained in the data of this report. All companies were

told that present costs for NODLR systems were not acceptable to the Army. Each company was told that a figure of 35K for production systems might be acceptable. The average value for systems in production by our analysis was around 39K. It is not clear whether this average value is coincidental with the number 35K or whether the suggestion of 35K somehow forced the average value of 39K. Put another way, would the suggestion of 20K force the average value to a value near 25K? This could be a systematic error.

c. Plan

Since the decision was made to go to the companies themselves, the following plan was used:

1. Introduce companies to NVL's Risk Analysis Program.

   a. Explain aims of program and ground rules

   b. Explain benefits to the government

   c. Explain benefits to the company

   d. Give companies time to consider the program

2. Visit companies to obtain data.

   a. Use a questionnaire and on-site survey

3. Evaluate the results

   a. Construct company network and calculate probability curves.

   b. Show results to each company for agreement on network, for verification of inputs, and for possible changes.

   c. Recalculate networks with changes

   d. Identify problem areas

4. Verify the data

   a. Evaluate previous contracts of each company

      1. With NVL

      2. With the Air Force and Navy

   b. Evaluate opinions from neutral sources

   c. Revisit companies to discuss "obvious errors".

5. Present to management

d. <u>Implementation of Plan</u>

Since cost is of primary concern for this IR system, and the money for Phase II was not committed irrevocable to Glorious Systems, the following approach to the infrared companies (except Glorious) was considered reasonable. Assume that NVL has x amount of dollars to buy from 1 to 20 NODLR type systems. The ground rules are:

1. The system could be of any design as long as it meets certain performance specifications.

2. The specifications would be based on the Glorious contract.

3. The delivery of the systems would be based on the Glorious contract.

4. Cost of the 1st system was not as important as cost for system number 20. Cost is the most important parameter. Time is of secondary importance in that small time slippages are tolerable. It should be noted that the Glorious time schedule was one of the ground rules. It was decided to ask as many infrared companies as possible to help us determine the probability curves associated with cost to the government for from 1 to 20 systems within the next year. Associated with these costs would be delivery times and the technical competence of each company. As an added payoff, the latest system approach of each company would be revealed to us rather painlessly for us and for them.

The approach was new. The big question was whether the companies would go along with the approach. Telephone conversations with companies indicated that no company would commit themselves until a full outline of the program was given to them. In the vernacular it amounted to, "what's in it for me." So, early in the analysis it became quite evident that the success of this approach depended on aggressive salesmanship. At the first contact with each company, it was made clear that our request was not asking for a proposal and was not to be considered in any way as asking for a proposal. We were asking the companies to assist us with their professional experience. Free and open technical discussions between individual companies and the government were highly desirable. No company would be held legally responsible in any way to any answers given to the government. However, since risk analysis would be done on future NVL systems, it was in the company's interest to be reasonably accurate. The method of gathering information would be of administering a questionnaire to company group leaders. We would ask to be shown the various facilities of each group. The most sensitive of all areas, i.e. cost, would be handled in any way that they choose. Each was told that they would be compared with all other companies.

e. <u>The Carrot</u>

For those companies that assist us we would offer the following:

1. Their system ideas would flow through a direct pipeline to the top management of NVL. (Exposure or advertisement)

2. Each company would have the opportunity to review our model of their system <u>before</u> it is presented to management. Our model of them and our computer outputs of only their model would be given to them.

3. An industrial "average cost" and time distributions would be given to each company.

4. There would be a probability greater than zero that one or more of the companies could get part or all of the contract for Phase II.

This explanation with a list of the ground rules and essential specifications was given to all but one company by the same government representative. This one company did not assist in the program any way. Each company got the same sales talk. Each company was asked to think about the problem and let us know whether or not they were interested.

f. <u>Company Response</u>

All companies, known to be making real time IR imaging systems, were contacted. Six out of eight decided to assist us. One non-assisting company stated that they were interested in assisting in future programs but due to existing company programs and manpower requirements they could not help at this time. The other non-assisting company was Glorious Systems. This is understandable, since Glorious Systems had the NODLR contract. Nevertheless, they showed interest in future NVL programs.

4. <u>Data Collection</u>

The data was collected in two categories, system hardware and program software. Systems hardware refers to those times and costs associated with actual hardware development. The data began with the initiation of the contract and ended with the acceptance test. System software was denoted as "x", meaning that it was a variable. The items contained in "x" are reliability studies, maintainability costs, data packages and other government requirements. Very detailed hardware costs were obtained for up to 20 systems.

5. <u>System Data</u>

Each company participating in our analysis was visited and data was collected by the questionnaire method. Detailed data were obtained which satisfied a net typical of Figure (1). The details of the data for the entire system is exemplified by the path of the scan head. Even H10 initiates the design of the scan head. Games were played to answer the following typical questions.

What is the expected time to complete the design?

What is the earliest time?

What is the latest time?

How many men are involved in the design?

What will be the cost to the government? (This cost is a non-recurring cost and usually a fixed cost).

Event H21 ends the design phase and initiated the fabrication phase. In addition to the questions above, the following questions were asked.

In the fabrication what will be the cost to the government for the 1st system? What part of this cost is non-recurring?

What will be the cost for system #2, #4, #8, #16 and #20?

There were many other questions asked, but these give the general idea of the depth of the questioning. The same type of information was obtained for testing, possible redesign, retesting and integrating into the total system.

After information was obtained for all paths in the net, the data was fed into a MATHNET (a computer network developed by Mathematica Inc., Princeton, N.J.) program. Outputs were obtained and analyzed. After all systems were analyzed, the outputs and nets were brought back to the companies. Each company saw only its own output and net. Each company was asked to comment on the net. Any changes were encouraged. Areas that seemed to be in error were pointed out. Each company verified the nets along with any changes on them. All said that they were satisfied that the nets were a true representation of their inputs. These final nets were used as our input data.

There was a reason for asking detailed information for up to 20 systems. There was general agreement from industry that more than 20 systems would begin a production phase. Details of time and costs for under 20 systems were readily available from the various companies. No interviewed companies had details on production, since no company had ever mass produced systems of this type. Information was obtained for systems up to 300, but these data were not verified by us, therefore, we were not convinced. They were simply company promises. These facts were taken into consideration in our rating scheme.

6. Rating

a. A Comparison Scheme

Setting up a comparison scheme for our analysis was difficult because all the companies involved were technically competent and presented reasonable systems. The data for comparison fell into two main categories, subjective and objective. But, it can be argued that even some of the objectives "facts" are slightly subjective. If the problem were simple, there would be an established procedure that one could look up in a text book. Nevertheless, the situation is not hopeless. So, at the risk of re-inventing the wheel, a rating scheme was boldly used in this analysis in an attempt to minimize subjectivity. Inputs into the rating were obtained from our company questionnaire and a questionnaire drawn up especially for our own NVL people, and various Army, Air Force and Navy laboratories. This questionnaire served many purposes, such as adding to our data bank, assisting in rating the

companies, obtaining other government labs' professional experience and providing some verification to the companies input. The rating scheme arbitrarily was selected to give each company a maximum of 100 points. The simplified rating scheme follows:

1. Basis of Rating

   a. Company's cost of system #20. (Company promise) 30 pts max.

   b. Company's cost of system #300. (Company promise) 10 pts max.

This information is obtained objectively from our analysis. We feel that our inputs are accurate enough to believe that these figures are objective. The reason for assigning only 10 points to mass production figures is that our mass production data is not as detailed as our 20 systems data.

2. Company's Technical Level

   a. Technical Level (Total of 30 points max.)

      1. reliability of system (subjective) (10 pts)

      2. company's facilities and technical support (6 pts)

      3. company's understanding of problems in recognition vs system trade-off (5 pts)

      4. company's detector competence level (4 pts)

      5. company's cooler competence level (2 pts)

      6. company's scanner competence level (2 pts)

The reason reliability is called subjective is because so few NODLR type systems have been built that reliability is difficult to d rmine.

3. State of a battery operated system (20 points max.)

   a. passed acceptance test (20 pts)

   b. ready for acceptance test (15 pts)

   c. working system but not hardened for environmental tests (10 pts)

   d. breadboard model (5 pts)

4. Performance on all previous contracts (batting average - 10 pts max.)

excellent	10
good	8
fair	4
poor	0

b. Actual Rating

An experiment on this rating scheme was done on several people who were familar with the NODLR study. Each person was asked to make an educated guess to the following: If he had to invest his own money to buy systems for the Army, which systems would he choose? He was asked to rate all the systems from the first to the last. The results were recorded. Then, the same people were asked to rate the systems using the rating scheme. The results of the scheme were put on a blackboard. Subjective elements of the scheme were discussed by all members. Some members altered their "grades" on certain sections of the scheme. The results of the rating scheme were averaged, and the systems were rated from the first to the last. The top three systems by this scheme agreed with the top three systems chosen by the educated guess method. The fact that the results of both methods agree does not mean that this rating scheme is a formula for "instant success," but the fact that they did agree gives some confidence to the rating scheme. In this case it seemed to work.

7. Conclusion

Based on the rating scheme a set of alternatives were presented for management's consideration. They will not be given in this paper since they are routine for operation's research people. The main point of the paper is that development costs were predicted with some confidence. After companies were given our computer outputs, they were asked how close the results represented their costs. Answers ranged from 80% to 90% accuracy. Of course, the ultimate test is to buy NODLR systems by the open procurement route where the companies sign their names to the bid.

NODLR DEVELOPMENT FLOW DIAGRAM

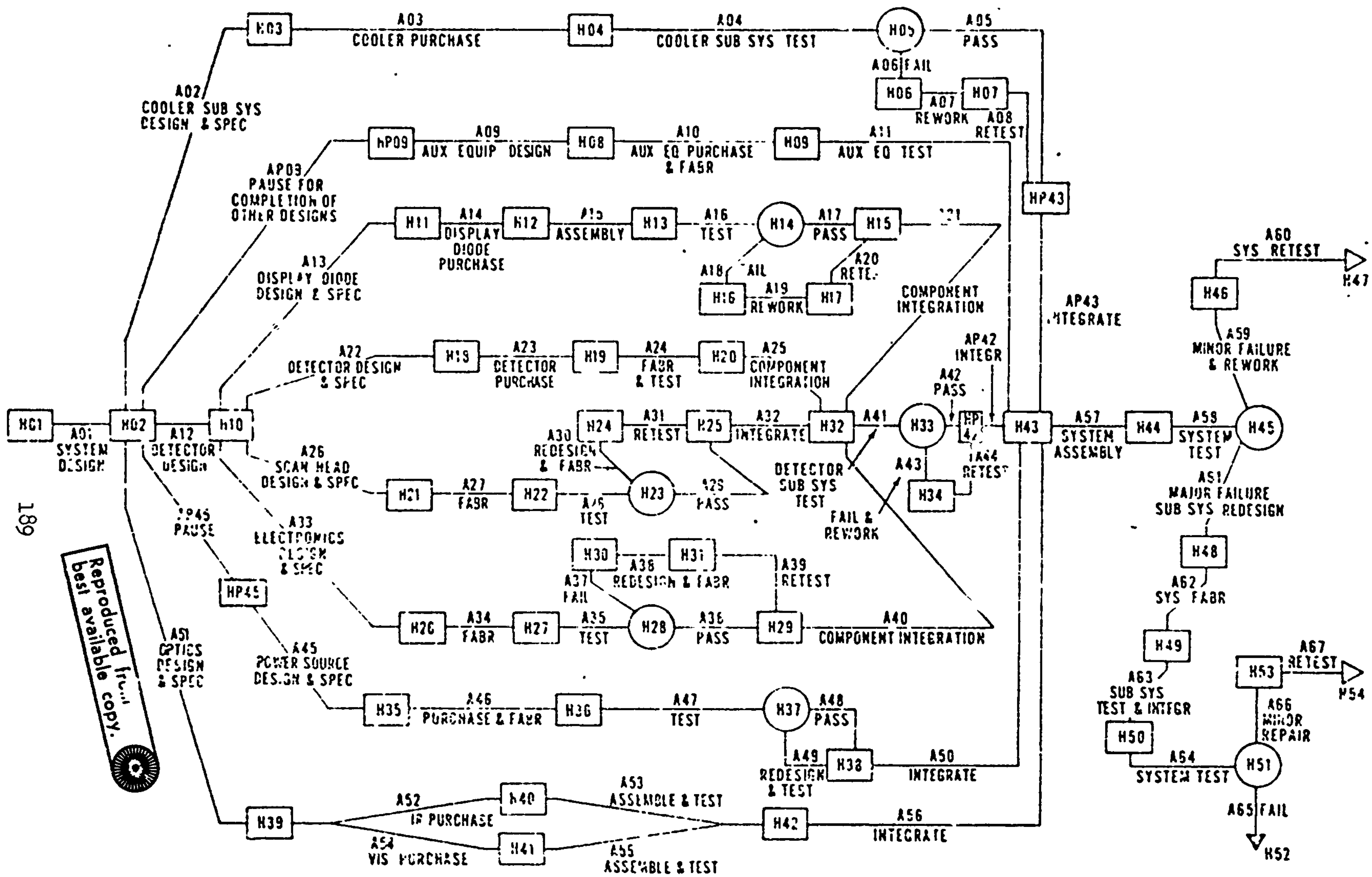


Figure 1

Reproduced from best available copy.

METHODS FOR EVALUATING THE EFFECTS OF EXPANDED BALLISTIC MISSILE THREATS ON DEFENSE DEPLOYMENT

Mathilde B. Sutow
William J. Douglas, Ph. D.

Keystone Computer Associates

1. Introduction

The determination of optimal defense configurations against ballistic missile threats poses many complicated geometric and kinematic problems in which it becomes difficult to isolate the key variables. The interaction between the parameters involved cannot, in general, be expressed in a closed form, and therefore lends itself to numerical solutions facilitated by the use of computer models. The purpose of this paper is to show how such models have been applied to study defense requirements against various threats in order to determine the options from which the defense can benefit most.

A computer model, MARC (Model for Area Coverage), has been developed together with associated pre- and post-processing algorithms which can be used to determine the area defended by an interceptor battery under a ballistic missile attack. The power and range requirements on the defense radar are computed and the best engagement is selected. The offense booster capability and the defense constraints may be widely varied. In subsequent sections of this paper, MARC is briefly discussed and its application to the comparison of several defense options is shown for ICBM and SLBM type threats. The examples selected are typical systems for Hardsite and National Value Area Defense.

2. An Area Coverage Model

MARC is a computerized model that calculates the interceptor (and radar) requirements for the defense of a given area against a ballistic missile attack. Given a system with constraints on some of the required parameters (for example, a maximum radar detection range), the section of the attacked area that may be defended subject to this limitation is computed.

The defended area is modelled as a grid of points against which several trajectories approaching from various directions are aimed. A spectrum of trajectories representative of the offense capability is considered. An interceptor battery at a given location with either a ground radar or a sensor not limited by ground clutter, is responsible for the defense of all the aimpoints. The re-entry vehicle (RV) must be intercepted outside the over-pressure damage volume which is computed from the weapon yield and the target hardness (1).

Preceding page blank

The keepout volume consists of the region enclosed by the over-pressure limit surface superposed upon a circle of given radius. This keepout volume can be placed at each aimpoint. This is done in the study of National Value Area Defense, where each aimpoint is considered to represent a city center. Another option employs a single keepout volume inside the aimpoint grid. In this case, the defense objective is to engage all RV's which attack the aimpoints and penetrate this keepout volume. Both examples are discussed in subsequent sections of this paper.*

The interceptor must be launched in time for its rendezvous with the RV, after the RV has been tracked for a certain period of time and its trajectory has been established. The point at which the RV is picked up by the radar and track has begun is referred to as the acquisition point. The position of the RV at the interceptor launch is referred to as the commit point. In the case of a ground radar, the points of acquisition, commit, and intercept must all be above the radar horizon (inside the clutter angle cone). The interceptor may not be launched until a given time after launch of the attacking missile. The point of intercept must be high enough to offer the defense the requested battle space. Battle space is defined by a given distance along the trajectory above the point of latest intercept.

Subject to all of the above constraints, the points of acquisition, commit, and intercept are computed for each aimpoint so that the radar range is minimized. Using this technique, the maximum radar range needed for the defense of an aimpoint against all the trajectories threatening it is found. The interceptor footprint includes all the aimpoints for which a set of acquisition, commit, and intercept points is computed for every attacking trajectory. The radar/interceptor footprint includes all aimpoints that require for their defense, a radar range less than or equal to the detection range of the radar under investigation.

In the study of the Hardsite Radar Defense, the requirements for the defense of a radar of a given hardness are the maxima over the requirements for the defense against each trajectory threatening the radar overpressure damage volume. Figure 1 illustrates the geometry employed in studying Hardsite Radar Defense.

The threat may be defined in two ways. One is by a set of trajectories of a given ground range, re-entry angle, ballistic coefficient and approach angle. This method is valid for an ICBM type threat where the trajectory ground range and approach angle do not vary significantly

*Other models of the keepout volume are available (2), but are not described here.

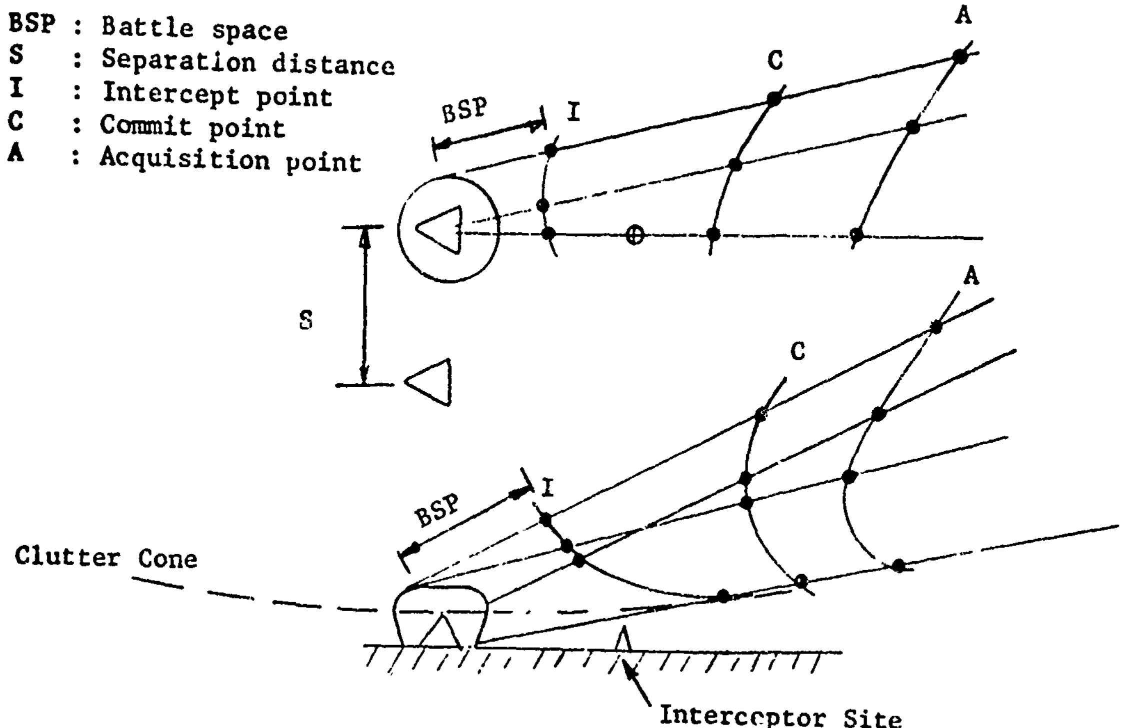


Figure 1

Geometry For Hardsite Radar Netting

from aimpoint to aimpoint. The second method is by trajectories launched from a set of points a given distance away from the coast. This is the defense enforceable "keep-off" distance. The trajectories are defined by given ballistic coefficients and booster burn-out characteristics (burn-out velocity and position vectors). This option is used in the study of SLBM type threats. Figures 2 and 3 illustrate some of the geometric parameters which are considered in studying ICBM and SLBM defense. The RV trajectories are Keplerian. The projectile is considered as a point mass moving in an altitude-dependent gravitational field about a non-rotating spherical earth. The drag forces are considered inside the atmosphere. The aerodynamic characteristics of the projectile are described by a constant ballistic coefficient. The atmosphere is considered isothermal with a density varying with altitude (3).

The interceptor performance characteristics are input to the model as a table of position versus flyout time.

The reader will find a detailed description of the model and its applicability in reference 2.

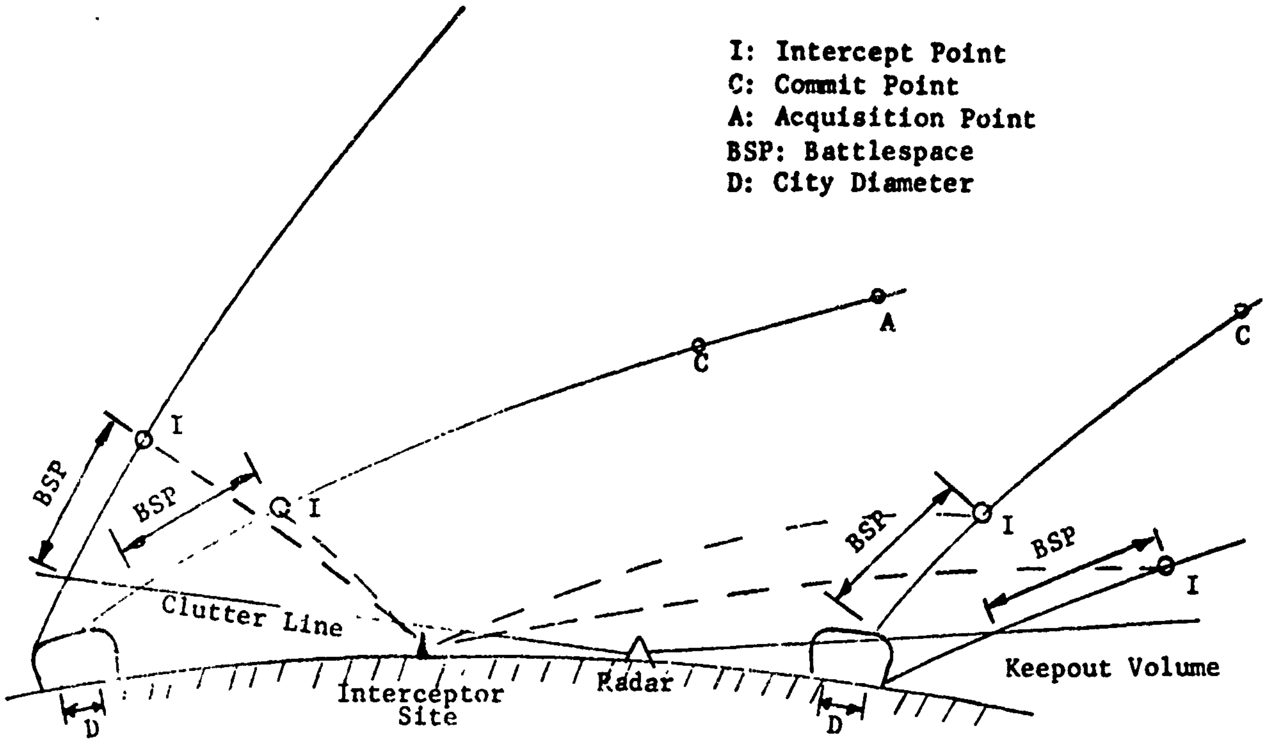


Figure 2

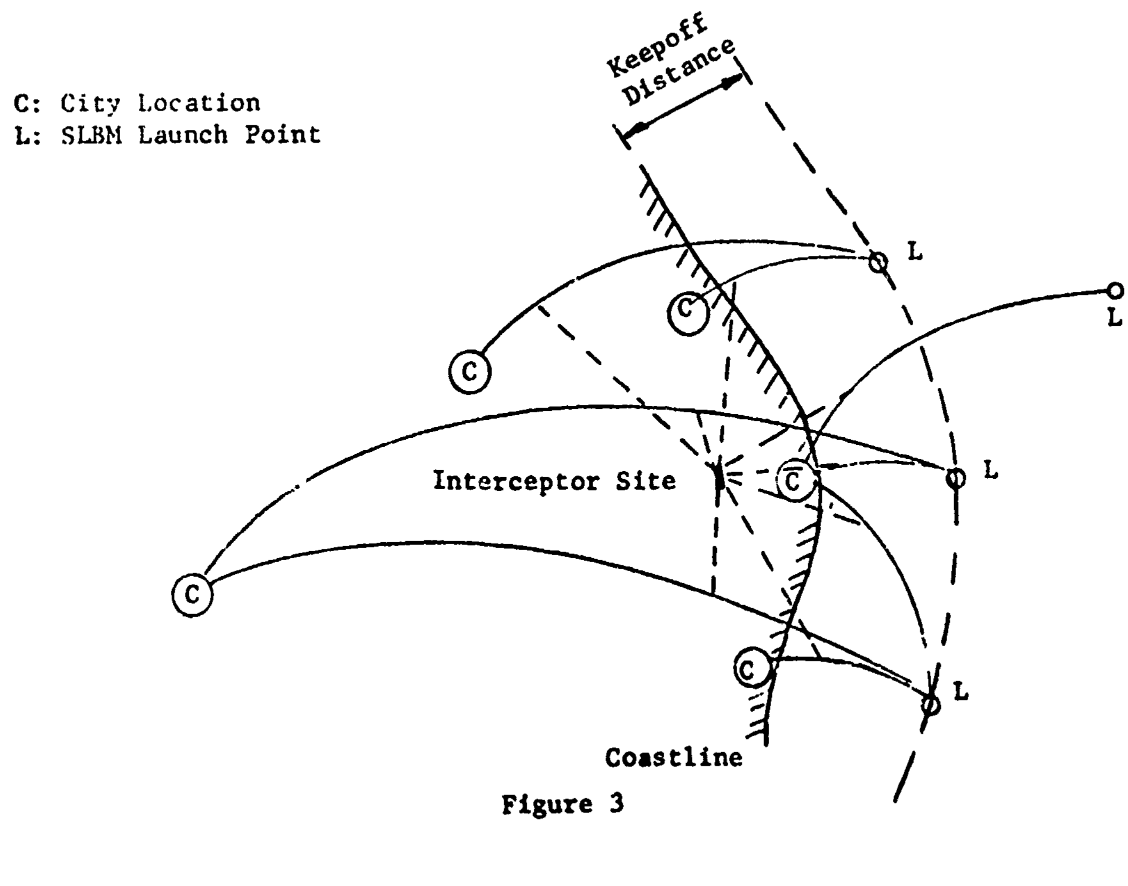


Figure 3

3. Application of MARC

3.1 Hardsite Defense-Radar Netting

In the study of hardsite defense systems, the defense of the radar is an important consideration. One defensive strategy employs a net of radars which must defend each other, in addition to a group of Minuteman silos. This mutual coverage poses a number of interesting constraints on the defense.

One problem which has been considered is to determine the effects on radar range, commit altitude, and radar power requirements for mutual (netted) defense as a function of radar hardness, weapon yield, net size (radar spacing) and battle space. The defense selects the interceptor site which minimizes required detection range. Using MARC, the radar net is evaluated by moving the interceptor site and determining the ranges and commit altitudes for each radar in defense of itself and all other radars in the net.

The outputs of MARC are plotted in various forms to allow one to study the offense/defense tradeoffs. Some examples are shown in Figure 4.

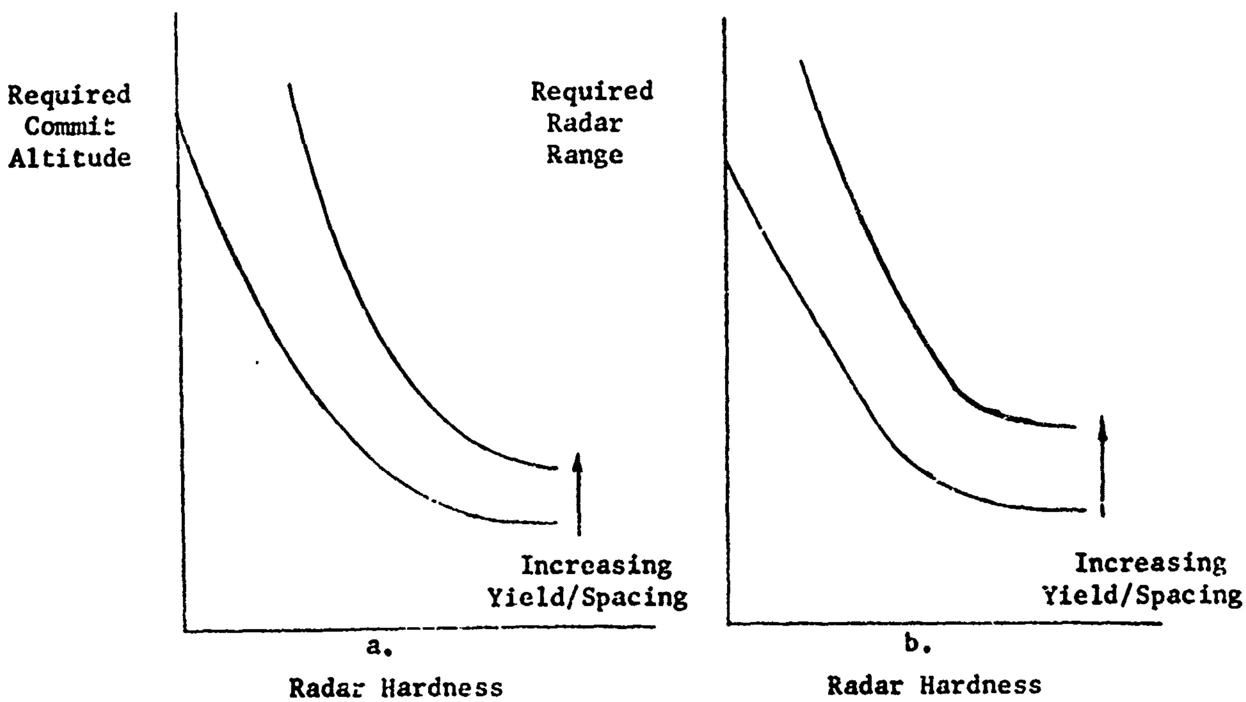


Figure 4

Figure 4 a., b. illustrate the variation of required commit altitude H_C and acquisition range R_{AC} for the mutual defense of radars. In each case, the offense weapon yield is varied and a corresponding spacing between radars is used.

Figure 5 a. shows the increased radar range requirements which result when the defense attempts to insure itself additional battle space. Figure 5 b. shows the radar power requirements for variations in yield, hardness, and radar spacing.

These curves can be used along with other defense limitations. For example, if the offensive decoys are effective to an altitude H_D (Figure 4 a.), then the corresponding hardness requirements are bounded by this value of altitude. Any decrease in hardening of the radar will require committing interceptors prior to discrimination.

Another example is shown by examining Figure 5 b. If the maximum radar hardness which can be achieved is defined by h_m and the power by P_m, then the range of applicability of the curves is limited by the dashed lines and another set of results is plotted in Figure 6.

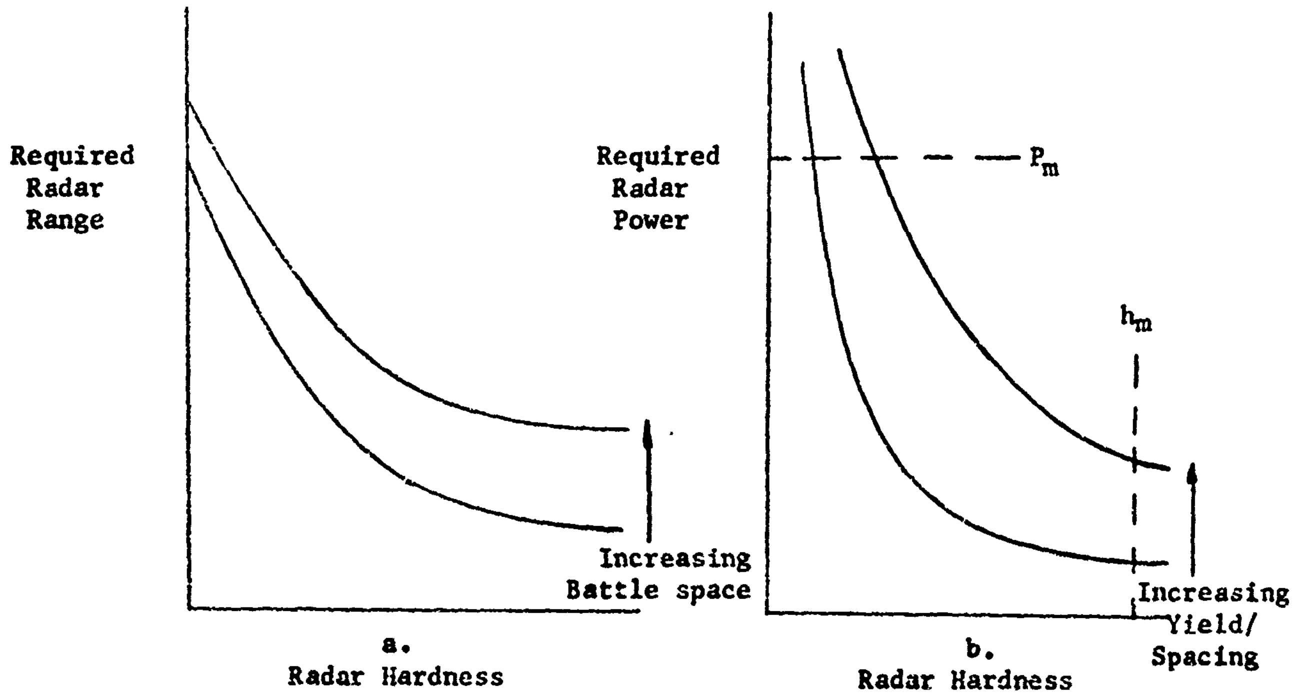


Figure 5

Figure 6 a. shows the variation of power required by the defensive radar vs. yield (and scaled spacing) of the offensive weapon for allowable radar hardness values. Figure 6 b. shows the variation of hardness required vs. yield for parametric values of radar power. One can conclude that hardening beyond certain values is not desirable, or that hardening produces severe penalties to the offense, depending upon the actual numerical values which result in Figures such as 6 a. and b.

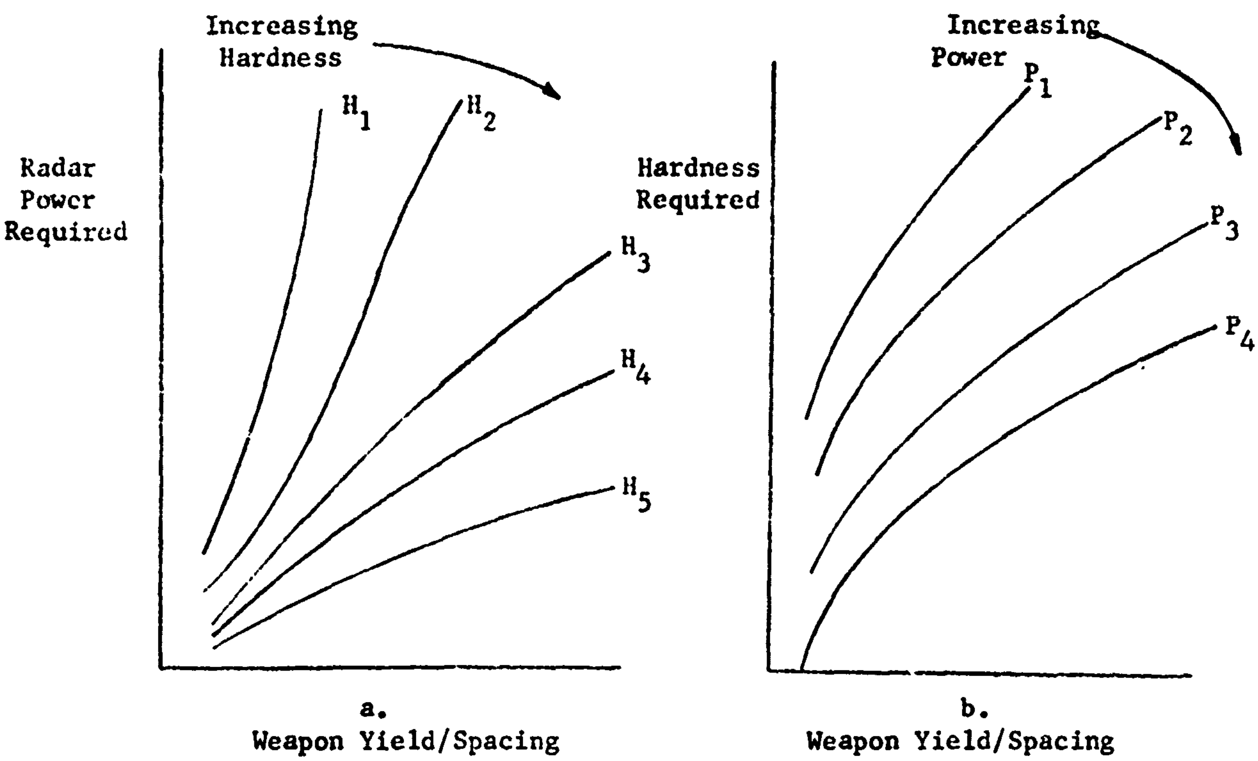


Figure 6

3.2 Area Defense Against ICBM Threats

In developing area defenses against ICBM's, the defense systems analyst is faced with a wide range of problems. These problems often deal with the area coverage which can be achieved by a radar with given viewing capability. MARC has been used to study a number of variations of the coverage problem. Figures 7 a.-d. show some of the important effects on area coverage.

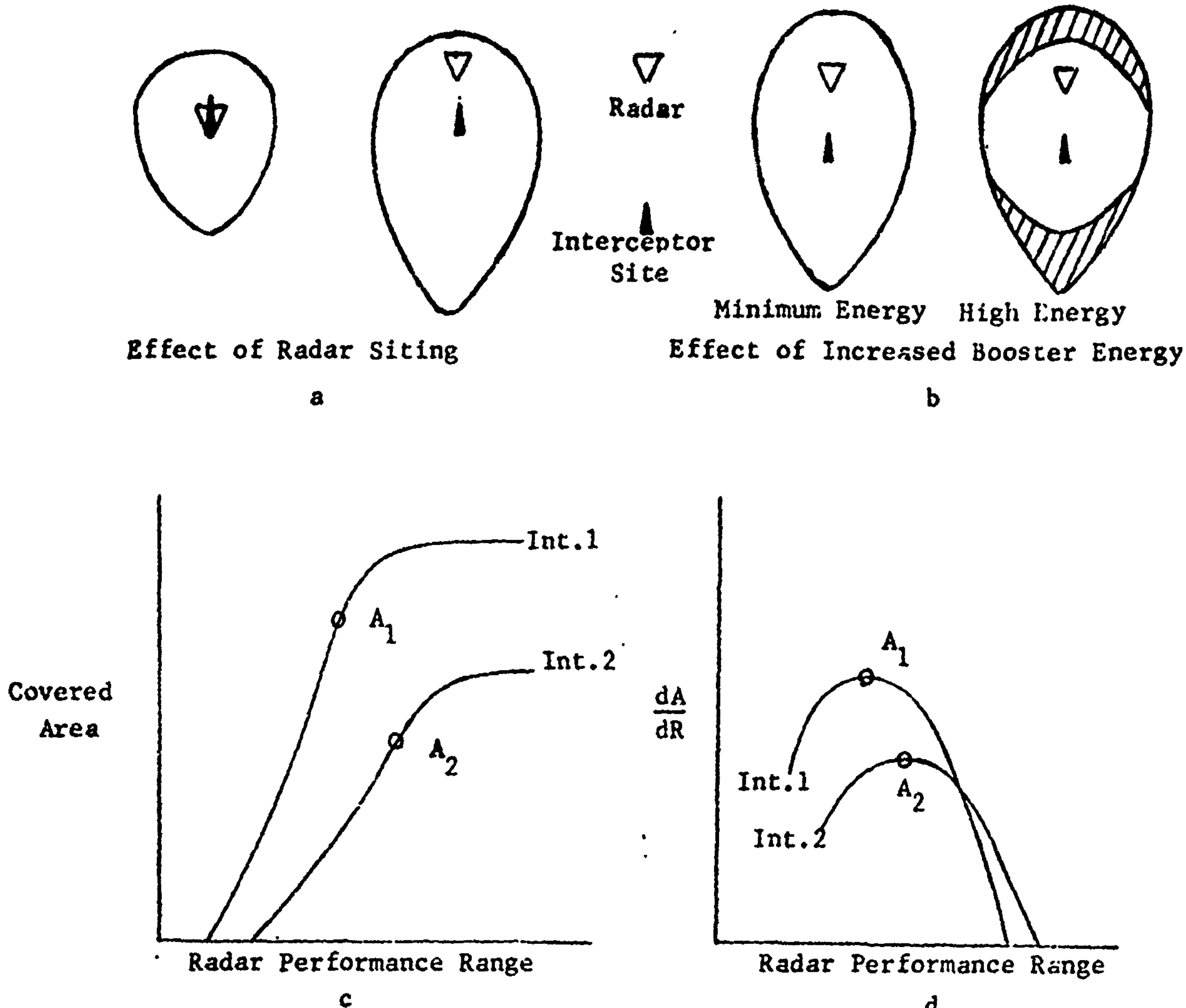


Figure 7

Figure 7 a. represents the effect of optimally siting the interceptor relative to the radar in order to achieve maximum coverage for a given viewing volume. The co-located system is compared with a forward radar system which is designed to take advantage of these effects. For trajectories targeted behind the radar, the acquisition range is shorter and the defense can balance acquisition and intercept range. Trajectories targeted forward of the radar become more visible, whereas they are not seen early enough for engagement in a co-located system. In addition, the decreased clutter for these trajectories allows for later intercept and consequently less interceptor performance (shorter flight time). Figure 7 b. illustrates the loss in coverage which results when the offense is given additional booster capability, which is represented by a wider range of re-entry angle limits. Forward coverage is decreased due to depressed trajectories, while rear coverage is decreased due to the lofting of trajectories.

Figure 7 c. shows the covered area as a function of radar range. If the radar is given more power and consequently a greater performance range, the interceptor is again sited to maximize coverage. The two curves shown by Figure 7 c. illustrate the effect of an increased performance interceptor (Int. 1) over a slower interceptor (Int. 2). In addition to being faster, Interceptor 1 has a greater volume of performance as is seen by the greater maximum coverage. Also, the figure shows that the rate of increase of coverage with radar range maximizes at points A_1 and A_2. This effect is shown in Figure 7 d. by a plot of the slope of the Area Covered vs. Radar Range. This figure shows the rate of change of coverage with radar range, and the maximum which results. By introducing a population density and by relating the footprints to U.S. population centers, the area covered has been converted to population to study the equivalent population coverage tradeoffs.

3.3 Area Defense Against SLBM Threats

The threat of submarine-launched missiles poses a number of new and interesting problems which clearly demonstrate the need of a model with a great deal of flexibility to show the effect of every parameter. Studying this problem one must be skeptical of simplifying assumptions commonly made for standard defense systems. The various offense parameters spread through wider ranges which are very much dependent on geography. For example, approach corridors up to 360^o are sometimes encountered, while the trajectory ground ranges can vary for different approach directions by an order of magnitude.

The effects of a number of key variables in the SLBM threat are illustrated by some examples. In Figure 8 an interceptor battery is sited in the center of the continental United States. The interceptor footprint with a radar optimally sited to maximize the area covered against an ICBM threat is shown together with its limited approach corridor. SLBM's, however, can be launched from a variety of sites which increases the approach corridor significantly. An interceptor battery with a co-located radar has a much smaller defended region. This is due to the limited visibility of RV trajectories which are launched and impact forward of the radar.

In the case of coastal batteries the short range of the RV trajectories becomes the most prohibitive factor. The defense is not given enough time to launch an interceptor and engage the RV before it reaches the damage volume of its target. Aimpoints that are further away from the launch site are easier to cover as shown by the shape of the interceptor/radar footprint in Figure 9. The defense can take advantage of this effect if it can force an increase in the distance between the launch sites and the defended region. The effect of different "keep-off" distances which can be enforced by ASW defenses is seen in Figure 9. The variation of the size of the footprint is shown for a coastal battery.

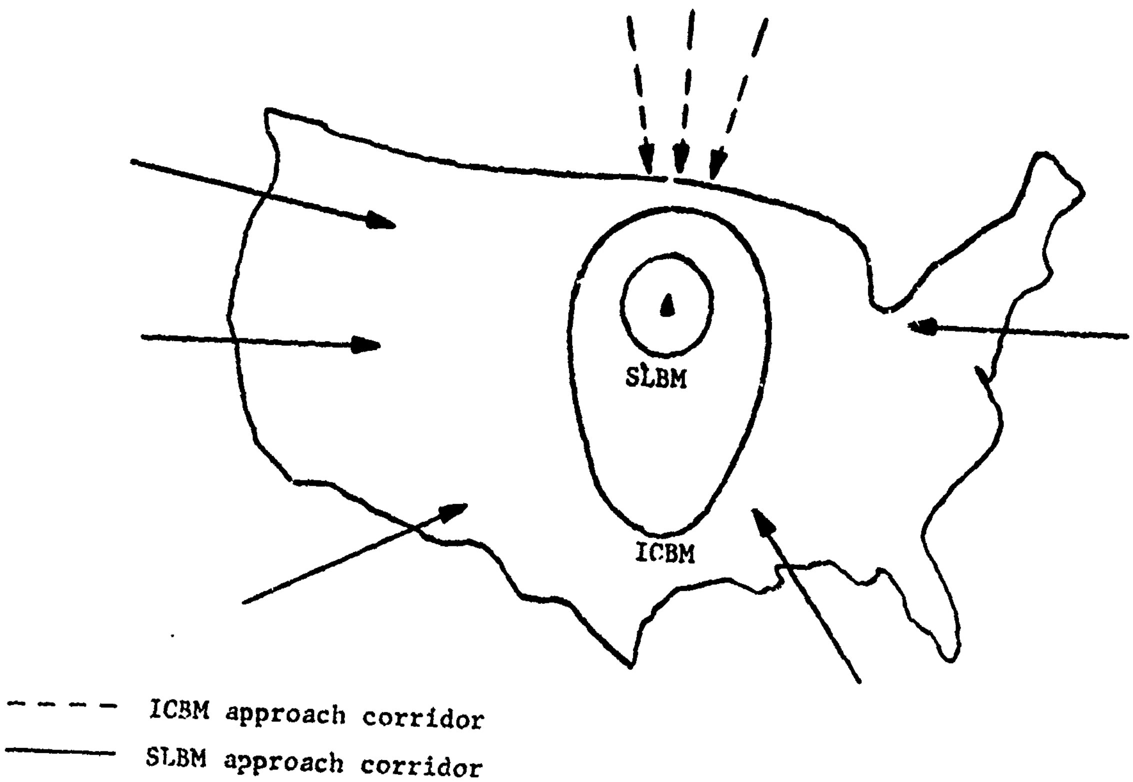


Figure 8

Effect of SLBM and ICBM Approach Corridors

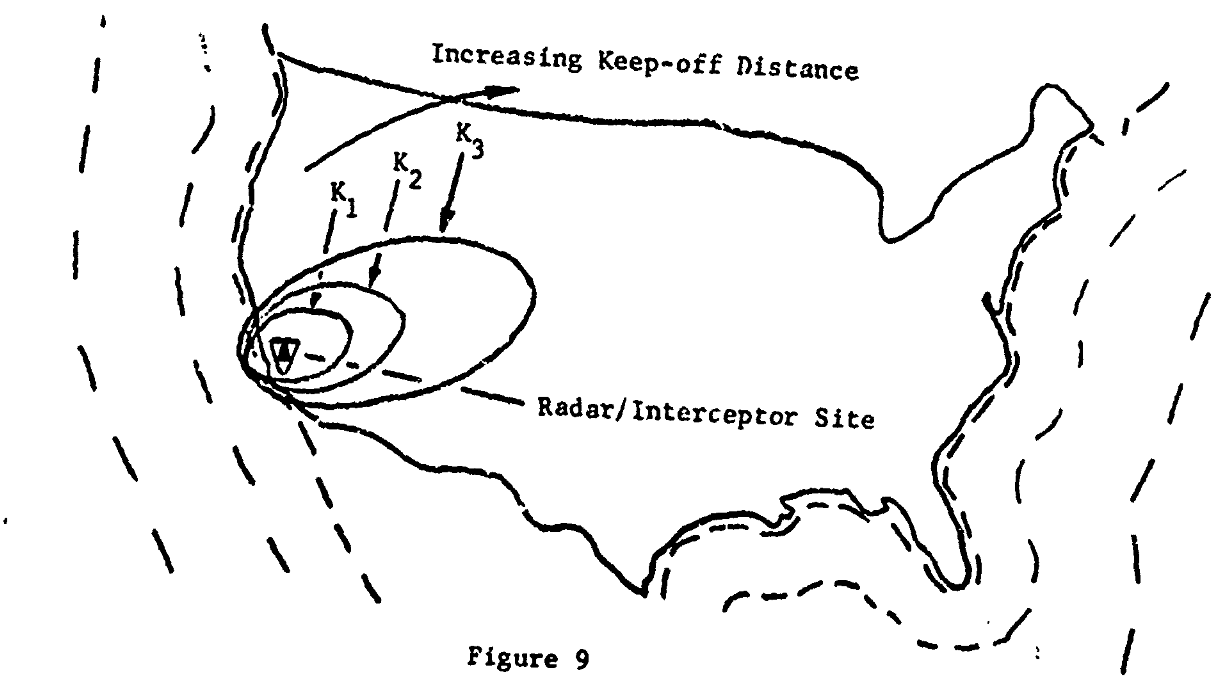


Figure 9

Effect of Keep-off Distance

To allow for an earlier interceptor launch, the RV trajectory must become as visible to the ground radar as possible. If the effect of ground clutter can be decreased, a larger footprint results as is shown in Figure 10. Making the trajectory completely visible by using an airborne radar or a satellite, the interceptor can be committed shortly after the booster burn-out or even before it. Thus, additional coverage can be obtained. The footprint which results by employing such a system is shown in Figure 10, also.

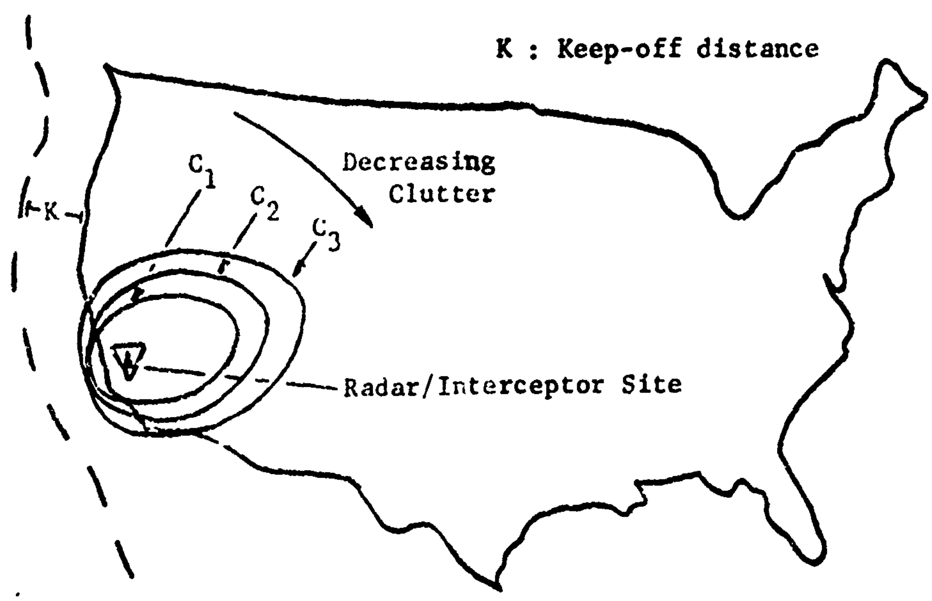


Figure 10

Effect of Decreased Ground Clutter

Finally, Figure 11 shows how a faster interceptor can be employed to win the race with the RV.

The gain in coverage which is obtained by enforcing different keep-off distances can be seen in Figure 12. An increased keep-off distance can ensure the defense a better coverage, while an improved interceptor can compensate for the cost-limited ASW defenses. Such data can support the evaluation of the risks, tradeoffs, and cost of alternative defense systems.

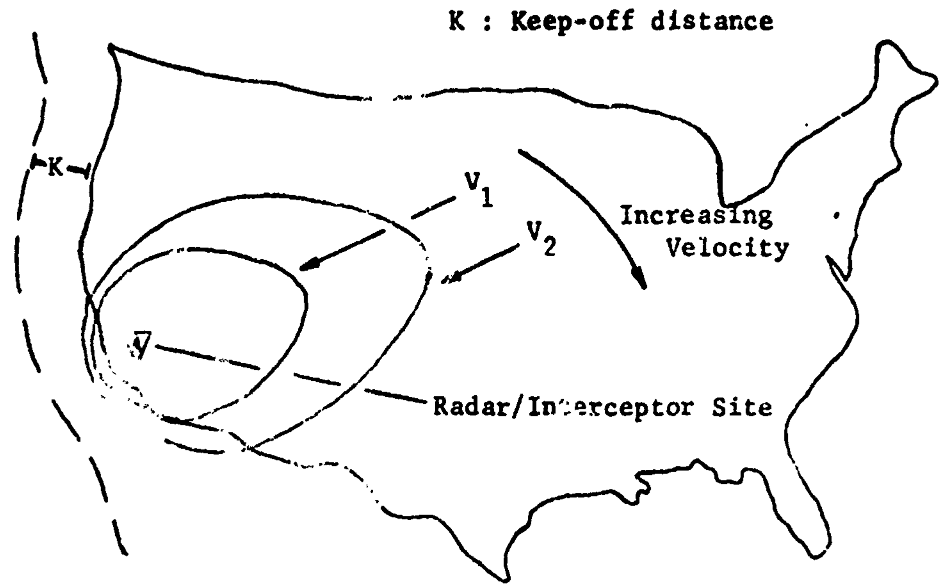


Figure 11

Effect of Interceptor Performance Characteristics

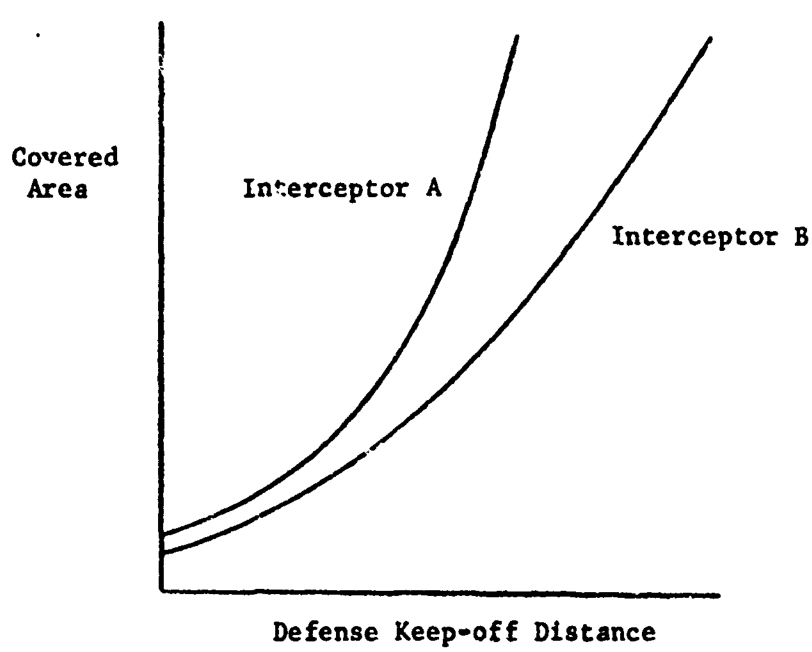


Figure 12

4. Summary

In studying BMD systems, optimization is a very elusive concept. The large number of choices available to the defense systems planner makes it difficult to design a system which is near optimal over all the variables. Consequently, the analysis of such systems is greatly enhanced by computer models which have many degrees of freedom to study the offense and defense parameters. Such models provide insight and direction to the analyst, and are effective and flexible tools for handling a wide range of problems.

A model called MARC has been used to study both Hardsite and Area Defense against ICBM and SLBM attacks. This model allows for variation of a number of key offense variables including ballistic coefficient, trajectory profile, weapon yield, and approach corridor; and a number of defense variables including interceptor performance, radar and interceptor locations, hardness, battle space, radar range/power, and commit/acquisition altitude.

Three problems have been addressed in this paper. These are described as follows:

- <u>Hardsite radar defense against ICBM's</u>

  In this problem, the radar netting requirements are established in terms of range, power and hardness as a function of offense weapon yield.

- <u>Area Defense Against ICBM's</u>

  This study is concerned with the effects of radar/interceptor siting and performance, and booster energy. The variation of area coverage with radar range and increase in covered area per unit increase in range are shown in a set of curves.

- <u>Area Defense Against SLBM's</u>

  Comparison of the footprint of an interceptor battery for an SLBM and an ICBM threat is made. Examples of various defense enforceable options such as an increase of the "keep-off" distance, the use of sensors not limited by the ground clutter and of interceptors with improved performance characteristics are illustrated.

In this paper an attempt was made to show the tradeoffs which can be studied with the aid of MARC. The results are fully parametric and quantitative conclusions are not drawn.

ACKNOWLEDGMENTS

The authors would like to thank Miss Marilyn Fleming for her significant contributions to the work described in this paper. Also, they would like to acknowledge the help of Miss Sally Hartzell and Mr. Stephen Schechner in the programming of the models. Finally, the authors express their appreciation to Mr. Willard Perry for his helpful advice.

LIST OF REFERENCES

(1) Gladstone, Samuel, "The Effects of Nuclear Weapons", U. S. A. E. C., 1962, pp. 130-148.

(2) Sutow, Mathilde B., "Modelling of an Area Defense System", to be published.

(3) Koelle, Heinz Hermann, "Handbook of Astronautical Engineering", McGraw Hill Book Co., 1961, Chapter 7.

Risk Analysis in the Acquisition of BMD Systems

Dr. Edward N. Dodson
General Research Corporation

1 INTRODUCTION

In recent years there has been increasing discussion and concern about Risk Analysis as part of the overall appraisal of military systems and equipment. The Secretary of Defense has called for:

> ". . .better risk evaluation of the uncertainties likely to be encountered in development. . . We have instructed the Military Departments that during concept formulation, they are to identify and analyze the areas of high technical risk. Where formal risk analysis shows that we are not ready for full-scale development, we will defer system development. . ."[1]

Similar views and recommendations have also been voiced by the Blue Ribbon Defense Panel (the "Fitzhugh Commission"), the Congress, and the President.

The US Army Advanced Ballistic Missile Defense Agency (ABMDA) directs the advanced development of BMD concepts and components including the identification of preferred system configurations, the definition of BMD "growth paths" responsive to changes in threat and technology, and the development of information for BMD decision-making. In support of these ABMDA missions, General Research Corporation has been conducting numerous studies, including several directly concerned with risk evaluation. The purpose of this paper is to outline the methods and general results of these studies.*

2 A GENERAL ANALYTIC FRAMEWORK

2.1 Background

Our studies of risk have evolved from an extensive background in parametric cost analysis. As part of our general resource-estimating and cost-effectiveness tasks we have also been concerned with estimates of lead times and schedules. However, in evaluating alternative weapon systems which might be acquired to perform some specified mission, single valued (or point) estimates of time and cost are not altogether sufficient because of the considerable differences among system alternatives in the likelihood of variations in estimates of time and cost.

---

*These studies have been fully documented in Refs. 2-4. A closely related study sponsored by OSD (Systems Analysis) is documented in Refs. 5-6.

In short, there can be significant differences among alternative weapons systems in the risk that development will not be complete in time to meet the threat, or that completion will not be feasible within the allotted funding.*

To complicate these widely recogniz points: management can commit resources to the achievement of a given design objective in a number of ways. A "crash" program with extra shifts, parallel development effort, concurrent production, etc., may permit reductions in time-to-completion, but this may require higher costs and may entail greater risk than a more "normal" pace. Briefly stated, there are alternative mixes of time, cost, and risk for each individual system corresponding to the alternative ways in which weapons acquisition programs can be organized and managed.

To account for these considerations we have developed a number of analytic techniques and procedures.

2.2 Network Analyses

As a general tool we have developed several network analysis procedures, the network format being a logical structure for analysis of the numerous, interrelated, one-at-a-time activities characteristic of R&D programs (as discussed subsequently, we have also adapted this format to analysis of later phases of the acquisition life-cycle). In form this representation is similar to PERT and other network techniques. We differ from the usual PERT analysis in two major respects. First, as others have noted, the PERT procedure computes the critical path on the basis of expected activity times--all subsequent calculations assume that the critical path remains along the same activities. In fact, in complex programs with a number of concurrent activities, there is a distinct probability that the critical path may shift from the most likely path. As documented in Ref. 7, the effect of the PERT assumption is to bias estimates of total completion time--the PERT estimates may be considerably less than the actual estimate should be. Thus, our approach was devised to include this critical path "switching" phenomenon.

The second point of difference with PERT and other network techniques is the failure of these approaches to fully account for the numerous interactions, or correlations, of times and costs among the various activities. One immediate consequence of interactions among time estimates is that the PERT calculation of total time variances is incorrect. Given correlations among the individual distributions of activity time, the variance of the sum along the critical path (i.e.,

---

*Risk, then, is defined in terms of likely variations of time-to-completion and cost. This facet of risk has been the major focus of our initial studies. As will be discussed, we have recently been looking at additional dimensions of risk which are especially pertinent in Advanced Development (as distinct from Engineering Development and later phases of the life cycle.

total project time) is greater than the sum of the individual variances (in statistical terms, one must add twice the sum of distinct covariances). Thus, the conventional PERT calculations (which do not account for covariances) err in their estimates of the expected variation in project completion time. Estimating this variation is crucial to proper assessment of program risk.

In practice, PERT-COST has been used primarily for purposes of cost control, and--to our knowledge--has not been used as a planning device in which interactions of time and cost were fully represented.

The various interactions of times and cost warrant amplification. Time-cost interactions arise from many factors, among which are the following:

a. As a given activity exceeds its estimated time to completion, additional expenditures are required to complete the work. These expenditures include both direct costs and overhead costs.

   In this simplest of interactions, direct costs scale with the work remaining, while overhead scales with the time required to complete the work. This simple relationship becomes more complicated, however, because:

b. An <u>imminent</u> slippage in schedule will generally become evident at some point within the scheduled time. Where adherence to the schedule has high priority, management will counter with extra personnel, equipment, etc., in order to meet the deadline, or to minimize the schedule overrun. Hence, activity time may be as originally planned, but costs may be in excess (although time is the real culprit). With the wide range of potential management responses, time and cost overruns can assume a great variety of values.

The preceding points have dealt with the relationships between time and cost for a given activity. There is also a variety of ways in which interactions take place <u>among</u> activities, for example:

c. Given a schedule slippage in one activity, there may be extra costs incurred by concurrent activities terminating in the same event. These costs could be associated with maintenance of technical personnel who will be used in succeeding activities, etc.

d. A schedule slippage in one activity may prompt accelerations in succeeding activities in order to reduce total system time or to avoid shifts in the critical path.

e. Given a fixed project budget, a cost overrun in one activity may prompt rescheduling in succeeding activities in order to reduce total cost (or, conceivably, the cost per unit time, e.g., in a given fiscal year).

In addition to these specific types of interactions, there are more general interrelationships, such as common problems of technology or common susceptibility to inflationary impacts, which can affect a great number of activities.

For purposes of this paper, the specific analytic procedures to account for all these considerations can only be summarized (Refs. 2-4 provide greater detail). Our approach includes Monte Carlo procedures together with a full statistical representation of intra- and inter-activity correlations, i.e., correlations between time and cost for each activity, and among these parameters for a number of activities.

One immediate problem that may come to the reader's mind is that of the availability of data with which to specify all the statistical relationships of time and cost. Needless to say, fully documented, objective measures of correlations and covariances are not available. However we assert that these relationships do exist, and ignoring them for lack of empirical data can lead to serious error. Ignoring the correlations among times and costs presumes that--in fact--they equal zero. In many instances this is much more unreasonable than a judgmental estimate of some particular value.*

To develop a better understanding of these issues, the first phase of our studies included a number of sensitivity analyses to establish which parameters are important and which can be reasonably ignored. Those found to have significant impact upon overall program estimates and upon the variations of time and cost are worth further investigation.

3 RESULTS FROM INITIAL STUDIES

Some of the basic outputs of our analysis of specific engineering development programs are illustrated in Figs. 1-4. In Fig. 1 the individual data points are from individual trials of the Monte Carlo process; the collective pattern defines a joint density function of cumulative cost and elapsed time through selected events, or milestones, such as completion of engineering design.** The data points can be projected to the individual axes, yielding the time and cost histograms of Figs. 2 and 3.

---

* There is a growing body of data which indicate the likely range of variation in aggregate time and cost estimates. Data sources include the Selected Acquisition Reports (SARs), US Army studies and earlier RAND studies.

** The contours of the bivariate density function can readily be drawn on Fig. 1 to include, say, 90% of the data points; then 80%, 70%, . . .

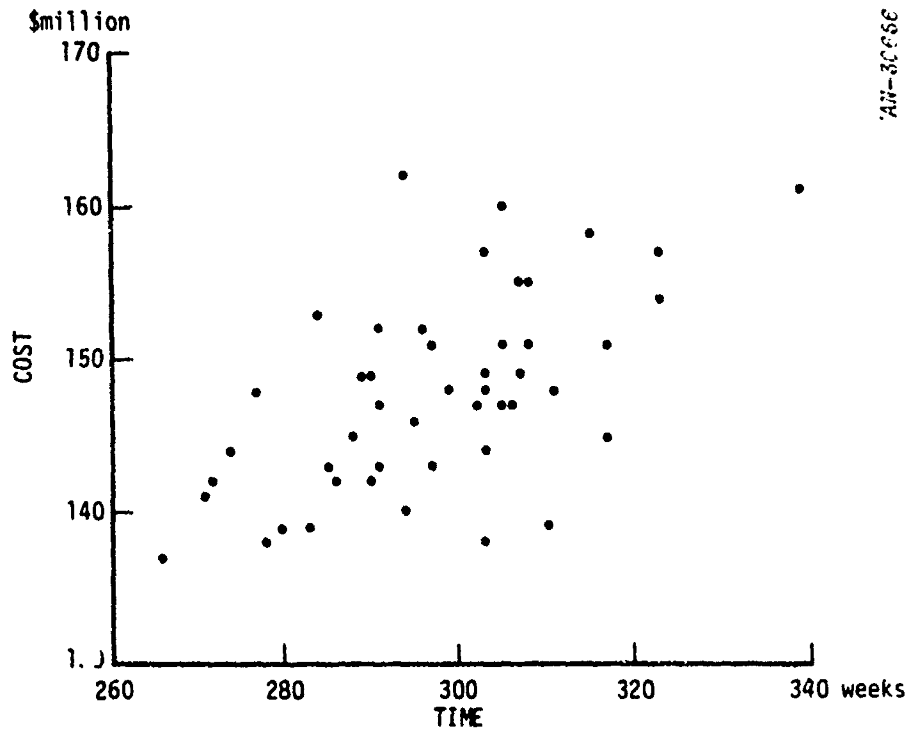


Figure 1. Time Versus Cost for Completion of Engineering Development

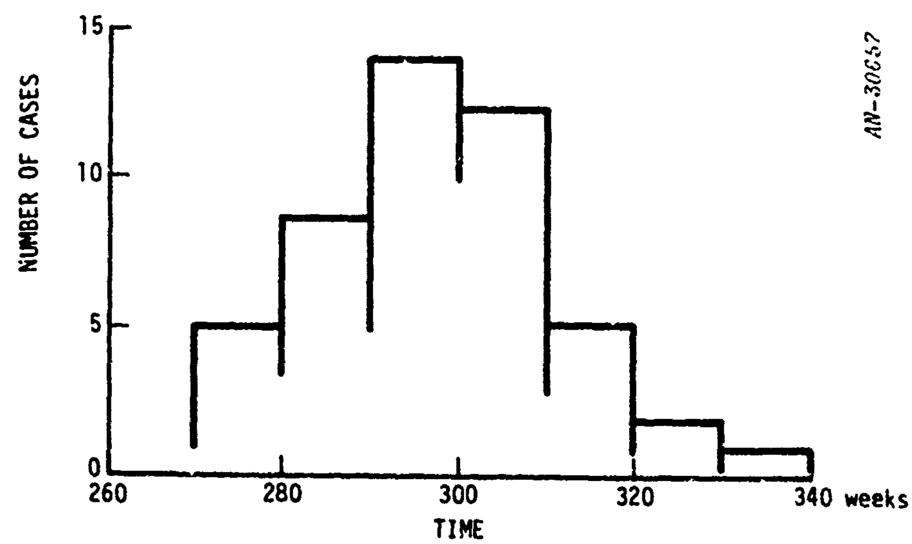


Figure 2. Time Versus Frequency

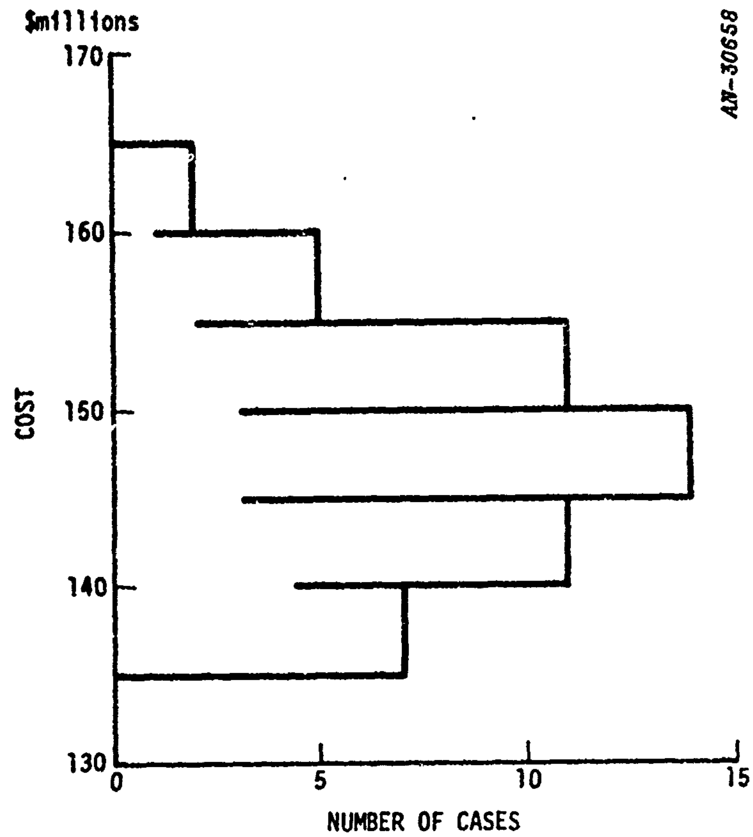


Figure 3. Frequency Versus Cost

These results--and comparable figures for other milestones--are summarized in an overall Time-Cost Summary. Also available is a printout of the "criticality" of each activity. As illustrated in Fig. 4, the criticality results for an individual activity are presented in terms of computed slack times (zero slack time indicating that the activity is on the critical path). The height of each bar may be interpreted as the probability (in percent) of the indicated slack time.

Our studies of actual development programs indicate that the phenomenon of critical path switching is fairly common, especially when the program has been planned for a rapid development pace.*

*Faced with the need for quick pace of development, management will naturally attempt to reduce the length of the critical path. Since individual activities generally cannot be "compressed" much, the more common expedient is to proceed concurrently with two (or more) activities that would otherwise be carried out in sequence. This adds another path through the network--one which will typically be near-critical. This means a greater likelihood of switching from the "expected" critical path.

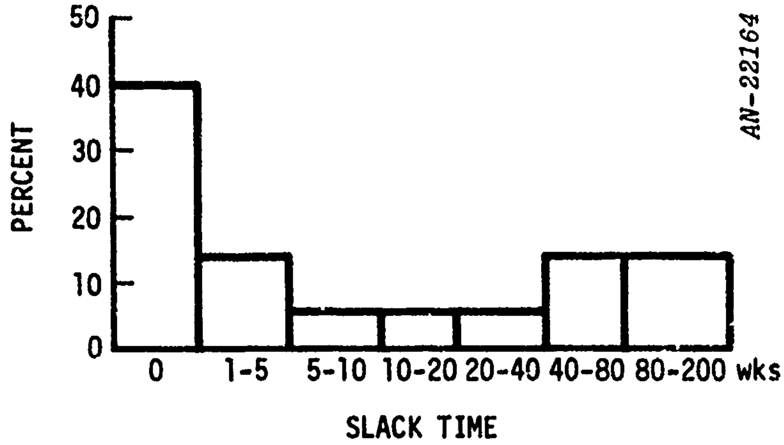


Figure 4. Frequency Distribution of Slack Time for a Particular Activity

---

Alternative Management Strategies. Figures 1-4 illustrate the output for one management strategy, i.e., for one particular plan (network) of activities, including a particular set of time and cost parameters. An alternative strategy means an alternative network; for example, Fig. 5 depicts part of a network for a hypothetical missile development program. In this example, a comparatively "radical" strategy is employed, in which rocket motor development is initiated prior to Contract Definition. (This strategy might be employed if the rocket motor were clearly the pacing item, and quick deployment of the system were considered urgent.) An alternative (and the more conventional) strategy would be to initiate rocket motor design at A102E, i.e., after Authorization to Proceed. Or, if the rocket motor were truly critical, a duplicate development contract could be awarded (and represented by a parallel branch in the network).

Initiating development of the rocket motor concurrent with Contract Definition (CD) entails some danger that the missile system will be so revised that the original motor specifications will no longer be applicable. Hence there may be a waste of expenditures and time in starting and scrapping one design, and initiating another. Comparisons among the alternative strategies can be made from the individual Output Summaries and time-cost histograms, and we have done this for a variety of alternative strategies.

Several noteworthy observations came to light in our initial studies. First, discussions with cost analysts indicated that point estimates of

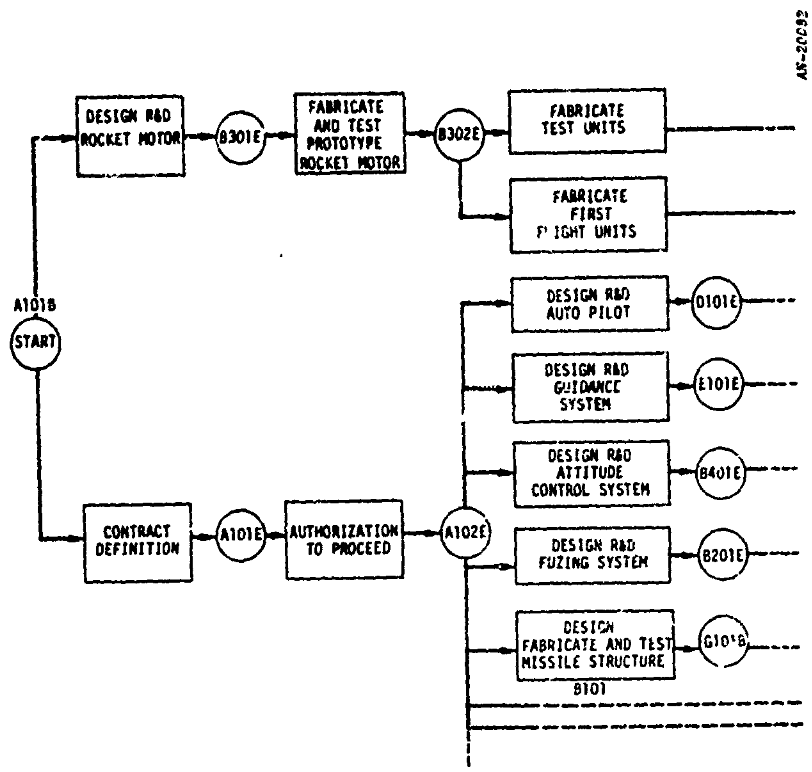


Figure 5. A Partial Network Diagram for Missile Development

time and cost could best be characterized as the most likely single-values (or modes) of the actual, underlying probability density functions. Moreover, these functions were almost invariably skewed to the right, i.e., there is greater likelihood of "actuals" exceeding these estimates than of being less. In our studies we have incorporated this skewness as an integral part of the network analyses. In statistical terms this says that the mean (or expected) value is greater than the mode. This difference can be of considerable importance, for it is the sum of mean values that is most significant in aggregating individual activity resources to determine total times and costs. Thus, the sum of point estimates, or modes, can be significantly less than the sum of means. In effect, this represents a source of bias--or underestimate--of total times and costs.* (For example in one study of an Engineering Development program, the sum of the modes was $188 million, while the expected value of total cost was $220 million.)

A second source of bias in estimating total program leadtime can arise if the critical path "switching" phenomenon is not taken into

*This underscores the need for a verification of aggregate cost estimates in carrying out detailed parametric cost analyses.

account. As others have also observed,[7] the PERT analytic procedure, which does not account for this switching, overstates the likelihood of lower values of total time. Thus, the expected value is actually greater.

As noted previously, we also carried out a number of sensitivity analyses in order to establish which judgmental parameters of the network are significant in terms of their impact upon estimates of overall times and costs. The results of these analyses may be summarized as follows:

- the particular shape of the individual time and cost probability distributions is not significant except as it affects the previously mentioned difference between modal value and mean value. The greater the difference, the greater is the "bias".

- correlations among the activities have significant impact upon total program risks, i.e., upon the likelihood of overruns in time or cost. As the extent and degree of interactivity correlations is increased, the range of variation in estimates of total times and costs becomes greater.*

- a key element affecting cost overruns is the management response to imminent schedule slippage. In actual practice there are often considerable extra expenditures in order to maintain delivery schedules (or to minimize schedule slippage). The nature of this cost-time function is, of course, difficult to ascertain, but sensitivity analyses of several likely functions indicate that the issue is significant to overall program results.

*This same point has been made in striking fashion in the literature of common stock portfolio selection (see, e.g., Ref. 8). These effects stem from the simple statistical rule that the variance of a sum of individual probability functions is the sum of individual variances plus twice the sum of distinct covariances. One conclusion to be drawn from this point is that a risk analysis should include a specific inquiry into economic or technical influences which impinge upon more than just a few program activities (e.g., perhaps a system has several components dependent upon some improvement in production techniques for sensor detector arrays, or upon some common resource such as software development experts. . .).

4 STUDIES OF EVOLVING BMD CAPABILITIES

In a second phase of study, we expanded the scope in several respects. First, we were concerned with the full life-cycle from advanced development through production, deployment, and activation. Second, rather than a single specific component, such as an interceptor missile, we wanted to include all the major components of a system, viz., interceptors, radars, forward sensors, etc. We also wanted to incorporate the various equipment changes and system modifications of an evolving defense capability to meet an evolving threat.

Our network representation was thus expanded to account for several component acquisition programs undertaken to meet a near-term threat. Concurrently, more advanced BMD components are proceeding through the acquisition life cycle so that they may be phased into operation to meet more advanced mid- and long-term threats.

The analytic results of these studies are similar in form to those illustrated previously, viz., they include probabilistic representations of times and costs, the effects of alternative management strategies and program plans, and identification of critical-path programs and available slack times. The slack times indicate how long particular program decisions can be deferred without affecting total system delivery schedules.

Toward a More Formal Representation of Risk. The ABMDA 1970 Indian Summer Study provided an opportunity to develop more explicit representations of risk in terms of specific design parameters. As we hypothesized in an earlier study[5,6] the likelihood of achieving specified R&D design objectives is a function of a number of factors, including available money and time, and the extent to which the "specs" exceed the existing state of the art. Drawing upon the considerable experience and knowledge of participants in the Indian Summer Study, the relationships between design parameters, resources, and likelihood of success (and the conversely related, risk) were summarized for a variety of components in the format illustrated in Fig. 6.* These relationships were used in determining "low to moderate" scheduling inputs for overall acquisition plans. Where these plans resulted in delivery dates which were too late to meet the threat, the comparative risk inherent in any "acceleration" could be viewed.

During the Indian Summer Study other facets of risk were also considered. For example, some system candidates were particularly sensitive to variations in enemy tactics and procedures. With other candidates were associated certain political hazards. . . In short, a variety of operational and political dimensions of risk were also evaluated.

*From Fig. 6: scheduling two years of Advanced Development to achieve a design objective of Y_2 would be characterized as medium risk. A less stringent objective, Y_1, would be low risk.

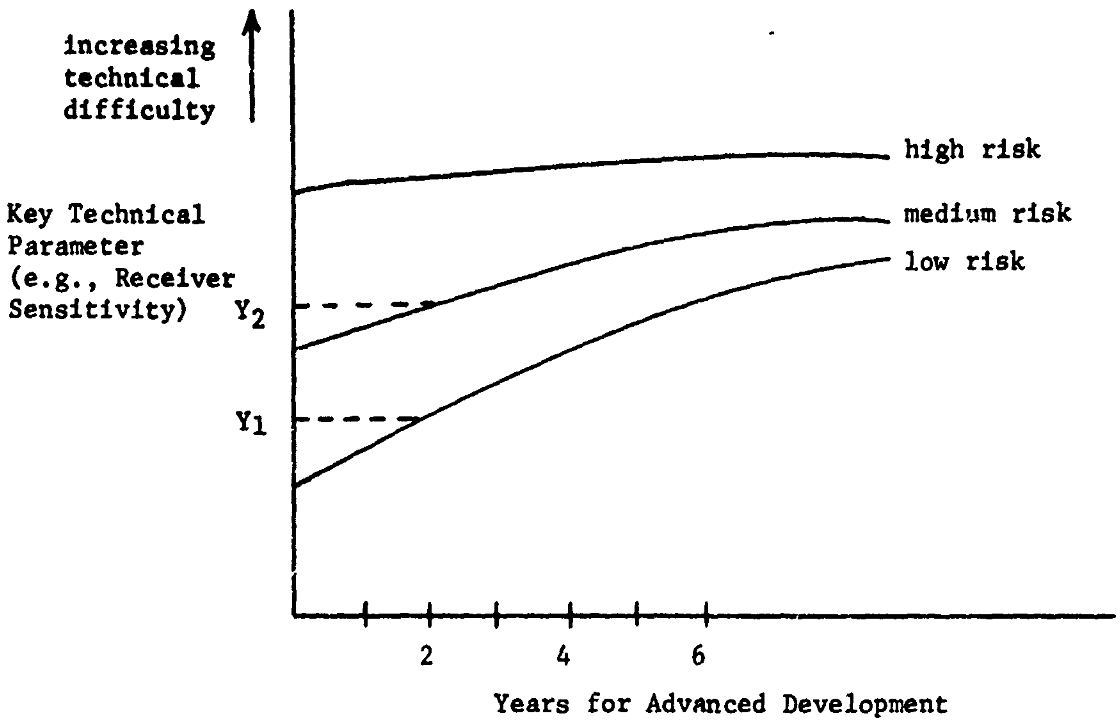


Figure 6. Parametric Risk Representation

On-going Efforts. Currently, our studies include refinements of the design-risk relationships illustrated in Fig. 6. We are also looking more closely at the nature of risk in Advanced Development (AD). Our earlier focus upon Engineering Development and later phases of the acquisition life-cycle was concerned with cost and time overruns--we did not consider the possibility that the design objective simply could not be achieved, regardless of available resources.* In Advanced Development, however, complete failure is a distinct possibility. We are attempting to establish how successive AD test milestones reduce that possibility, and what the consequences of failure may be. To illustrate this latter point (and referring to Fig. 6), the receiver sensitivity of Y_2 may correspond to an acquisition range of, say, 1000 miles. Failure to achieve that design objective and having to settle for Y_1 might represent an acquisition range of 600 miles. Thus the coverage of a planned defense system is reduced. In order to maintain some initial level of defense capability more sensors will have to be deployed. . . In short, the consequences of design "failures" can be related to costs for extra components. In other examples the defense may have to fall back to some lesser capability.

*By directive (DoD Instr 3200.9), the technology is "sufficiently in hand" in Engineering Development.

REFERENCES

1. Secretary Laird in statement before Joint Senate Armed Services and Appropriations Committee Hearings on FY 71 defense program and budget.

2. Time-Cost-Risk Relationships in Weapons System Development (U), Vols. I and II, General Research Corp. CR-3-144, July 1970 (SECRET).

3. E. N. Dodson, "Time-Cost-Risk Relationships in Weapons System Development," Proceedings, 5th Annual DoD Cost Research Symposium, May 1970.

4. Lead Time and Risk Assessment for Evolving Defense Capabilities (U), General Research Corp. CR-10-144, March 1970 (SECRET).

5. Studies in Resource Estimation for Development Programs (U), General Research Corp. CR-0365-1, May 1969 (CONFIDENTIAL)

6. E. N. Dodson, "State of the Art Advance and Resource Estimation for Development Projects," Proceedings, 4th Annual DoD Cost Research Symposium, May 1969.

7. R. Van Slyke, The Uses of Monte Carlo in PERT, The RAND Corporation RM-3367-PR, 1963.

8. H. Markowitz, Portfolio Selection, J. Wiley, 1959.

RISK ANALYSIS OF CASUALTIES

Dr. Daniel H. Newlon
OASD
Systems Analysis

How much should be spent to reduce future casualties?

This paper shows that:

. the present approach used by some cost-benefit studies cannot provide a satisfactory answer to this question because of incorrect methodology,

. an answer is needed in order to avoid waste of life and waste of resources,

. risk analysis can answer this question, but

. the accuracy of the answer would be improved if hostile fire pay depended on risk and volunteers were used for dangerous assignments instead of draftees.

I.

Ideally, a cost-benefit study determines the desirability of a project by adding the maximum each member of the community who would benefit from the adoption of the project would pay rather than forego the project to the minimum the remaining members would have to be paid to tolerate it. If the sum is positive, then the study can recommend adoption of the project without making interpersonal comparisons. Adopting the project and compensating the losers would make some better off with no one becoming worse off.

A variety of tools have evolved to enable the analyst to measure the compensation required by those who contribute such resources as capital, land, or labor to a project. If the project also involves the loss of life, similar measures could be used for casualties. John McClelland lists several examples of Army studies which cost casualties in the same way as equipment.

"In a recent cost-effectiveness study, pilot replacement costs for rotary wing aircraft amounted to $88,000, including $43,000, for the cost of burial, plus payments of Veterans Administration and social security benefits to survivors.

A weighted average of $3,500 per casualty was used in another study of alternate combat vehicle designs...A similar weighted estimate of $8,700 was recently developed which was based on a ratio of 3 to 1 killed to wounded among tank crews. The KIA estimate amounted to $19,000 and the wounded $5,300. Included in the $19,000 figure were the following cost elements:

Casket	$168
Clothing	46
Flag	6
Transportation	218
Escort	302
Interment	280
Three Headstones	78

This adds up to a little more than $1,000 for the burial..." 1/

These estimates represent a divorce of cost-benefit from community choice. A tank crew member would not choose to die even if assured that his burial costs would be covered and his family would receive survivor benefits. If the Army recruited someone with suicidal tendencies, society would not condone his use. In attaching a specific price to someone's life, be it $3,500, $19,000, or $50,000, the analyst is imposing his judgment on the person who loses his life.

II.

In a recent article E. J. Mishan suggested an alternative way of evaluating casualties.

"...the relevant sums to be subtracted from the benefit side are no longer those which compensate a specific number of persons for their certain death but are those sums which compensate each person in the community for the additional risk to which he is to be exposed." 2/

1/ John McClelland "Peacetime and Wartime Costs in Cost-Effectiveness Analysis" Economic Analysis and Military Resource Allocation, Office of Controller of Army, 1968, pp. 63-64.

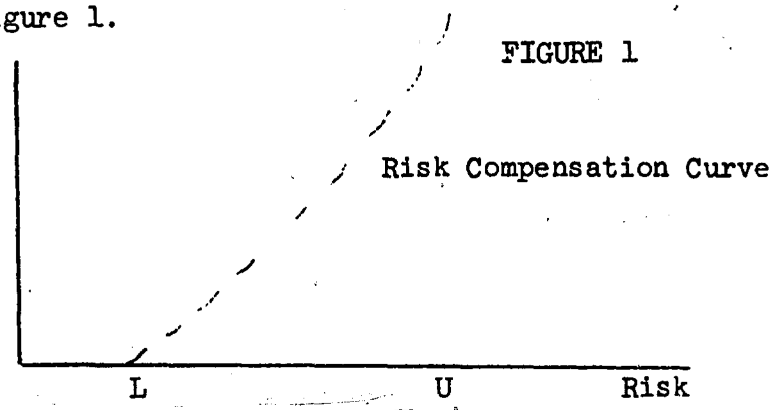
2/ E. J. Mishan "Evaluation of Life and Limb: A Theoretical Approach" Journal of Political Economy 79, No. 4 (July/August 1971): p. 694.

Mishan describes four types of risk depending on whether the risk is voluntary or involuntary, whether the indirect impact of the death is financial or psychic.

Returning to one of the studies mentioned in the last Section, the cost of different combat vehicle designs should include a compensation for the crew members that depends on the riskiness of the design. If the crew member volunteered, the risk would be of the first type; if he were drafted, the second. If the compensation were received by his family because of the risk of being deprived of his financial support, the risk would be the third type. If the compensation covered the risk of the family's psychic loss, then the risk would be the fourth type.

For a typical member of the crew, the compensation needed would be similar to Figure 1.

FIGURE 1



The crew member and his family do not have to be compensated if the risk falls below L because they are no longer concerned about the risk. But there are some risks that the crew member will not accept no matter how much compensation he is offered. The compensation required increases from the minimum sensible at an increasing rate, approaching the maximum acceptable risk asympotically.

III.

If the conventional assumptions used in microanalysis are made,[3] the Army will waste lives and resources unless

3/ For example, William J. Baumol, Economic Theory and Operations Analysis, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1965. pp. 362-65.

- the marginal cost of inducing a soldier to increase the probability of his becoming a casualty is equal to the marginal benefit, and

- the marginal cost of inducing a soldier to assume more risk is equal for all soldiers.

The first rule implies that the willingness of soldiers to risk their lives should influence the amount and nature of military service provided.

The second rule implies that the decision maker should be consistent. Lives are wasted if a large amount is spent to save lives and hence to reduce risk when a smaller amount invested in a different program would cause an even greater reduction in casualties.

The cost of inducing a soldier to assume more risk could be the additional training, or the improvement in the quantity and quality of his equipment that would reduce the risk to the former level. Therefore, implementation of the second rule will affect the distribution of equipment, and the amount and type of training received by soldiers.

Dangerous assignments should be shifted to soldiers with the lowest marginal cost to assuming risk. But assignments can become too specialized if the compensation for risk is not considered. Assume that two units of equal size and proficiency are under consideration for two different assignments. Both assignments will last a year but at the end of the year either unit would have 20% casualties in the first assignment and only 10% casualties in the second. The casualties are distributed evenly throughout the year. The distribution of soldier preferences for assuming risk is the same for both units with the average compensation schedule described by Table 1. If unit B received the more dangerous assignment and unit A the less dangerous assignment, both units would favor splitting the assignments and increasing the pay of soldiers in unit A on the average by an amount greater than $40 and decreasing the average pay for soldiers in unit B by an amount less than $150.

Table 1

AVERAGE COMPENSATION SCHEDULE

Risk (probability of KIA)	Compensation (per month)
5% or below	$ 0
10	10
15	50
20	200
25 or above	inf

In general the average amount of time spent in an assignment should be influenced by the benefits of risk sharing along with training costs, the advantages of experience, etc. A widening of differences in the risk of becoming a casualty should decrease the amount of time soldiers spend in the relatively more risky assignments.

IV.

Both rules could be incorporated into present cost-benefit studies. The risk premiums could be estimated and added onto the estimates of the prices of equipment, labor, etc. But even if one assumes accurate estimates from limited data, such a procedure would destroy the advantage of adopting this approach. The soldier who assumes more risk because of a change in assignment policy or the quality and quantity of weapons because of a cost-benefit study will be worse off unless he also receives <u>an increase in pay</u>.

In Section III the analyst cannot recommend assigning unit A to the dangerous assignment in place of unit B without a pay increase for the soldiers in unit A. Otherwise, he would arbitrarily decide that the soldiers in unit A deserved the more dangerous assignment despite their preference for a safer assignment. The new assignment policy would not conform to the ideal of making everyone better off.

But if pay accurately reflected the marginal cost of inducing soldiers to assume additional risk, soldiers would prefer the assignment policy recommended by cost-benefit studies. Soldiers in units A and B would prefer rotation with correspondingly higher pay for unit A and less pay for soldiers in unit B.

Such a pay scale would enable the cost-benefit analyst to contrast the compensation needed for additional labor, raw material, and equipment to reduce the risk of becoming a casualty with the cost of paying soldiers to assume the risk. The soldier who received the pay instead of the investment of better training and equipment would prefer the higher pay and correspondingly greater risk.

V.

Under a volunteer Army market forces would cause the pay to reflect the cost of inducing soldiers to assume additional risk If there were no administrative costs to special pay, a volunteer Army that did not discriminate between more and less risky jobs would be more expensive than a volunteer force that assigns soldiers according to their job choice and adjusts salaries until shortages are eliminated. The former policy pays rents to soldiers in the secure jobs in order to eliminate the shortfalls in the riskier jobs. Forcing soldiers to take the riskier jobs would not eliminate these rents because the decrease in the probability of a soldier's choosing his job would create shortages that would have to be eliminated by wage increases.

The draft distorts decisions in two ways. First, the observed cost will be less than the marginal cost of inducing a soldier to assume additional risk. Unless allowance is made for the difference, the marginal cost will be less than the marginal benefit. If observed costs are used, lives will be wasted.

Secondly, the marginal cost of a draftee assuming additional risk will tend to be less than the marginal cost of a volunteer. Assignment of volunteers to more dangerous jobs requires an increase in budget costs, while assignment of draftees can take place without changing the budget. If observed costs are minimized, the draftee will be placed in the more dangerous assignment regardless of his preference.

If the volunteer Army is politically infeasible during a war, the distortion caused by the draft could be minimized by assigning volunteers the dangerous jobs and adjusting pay so there were no shortfalls in these areas.

VI.

If the personnel and pay systems were restructured so that pay was associated with the job rather than the soldier and assignments were determined almost entirely by choice rather than command, soldiers who risk their lives would be compensated at less administrative cost than under the present personnel and pay systems with their proliferation of bonuses and options. 4/

4/ For a complete discussion of this proposal see Daniel Newlon "Some Algorithms for Reducing Job Dissatisfaction" unpublished manuscript presented at the Operations Research Society of America Symposium in April 1972.

But means for compensating soldiers who risk their lives already exists in hostile fire pay. Hostile fire pay rates could be adjusted, so that there was no shortage of volunteers for dangerous assignments. The rates that eliminated shortages would reveal the marginal cost of inducing soldiers to accept additional risk.

The proposal by the Army to pay a base rate for indirect support personnel in a combat zone, 125% of the base rate for direct support personnel, and 150% of the base rate for combat arms would seem to be a step in this direction. But an individual infantryman in a combat zone might have a lower probability of being killed than an individual cook. The rates are not sufficiently discriminating.

An alternative would be hostile fire pay based on deployment casualty ratios. The average deployment in a geographic subdivision of a combat zone could be divided into the number of soldiers killed in action or severely wounded and these ratios used to determine the probability of becoming a casualty. The larger the ratios, the greater the compensation received by soldiers in the area. The ratios could then be adjusted for MOS. The definition of the geographic subdivision would have to be large enough to minimize randon disturbances and small enough to reflect variations in probabilities within the combat zone.

In order to be administratively feasible, the soldiers would have to be paid after-the-fact. After the soldier completed his tour of duty he would file a voucher similar to the Temporary Duty form. The voucher would describe his assignments, i.e., units, time, MOS, during his tour of duty in the combat zone. The casualty deployment ratios for each unit and MOS in the combat zone could be tabulated monthly or even weekly based on the location of the units during the time period. By consulting the Tables, the soldiers appropriate ratios could be determined. The ratios could then be used to determine the lump sum hostile fire compensations.

The budgetary cost of such a system of hostile fire pay could be less than the present hostile fire pay. Assume that during the first of four years of a lengthy war 25% of the time an average soldier spends in the combat zone would be risky enough to justify hostile fire pay. During the second year the percentage decreases to 20%; the third year, 10%; and the last year, 5%. If hostile fire pay is a function of risk then a soldier who risked his life for one year in the combat zone could receive on the average $5,200 after his tour of duty had been completed at less budgetary cost than the present incentive of $65 per month for everyone in the combat zone. 5/

5/ These percentages are overestimates of the savings during the last four years of the Vietnam War. If one assumes in 1968 that all soldiers in the combat arms are continuously in combat and no one else risks his life and % decline in risk in 69 and 70 is the same % as the decline in the time the combat soldier risks his life the percentages drop from 23.4% in 1968 to 16.9% in 1969 and 7% in 1970. These gross estimates are used because the actual figures are classified.

The 1971 Quadrennial Review of Military Compensation criticizes the use of hostile fire pay to attract volunteers. "The historical purpose of Hostile Fire Pay is special recognition of the hardships and sacrifices endured by members in combat. This purpose differs from that of other special pays whose purpose is attraction and retention of volunteers to meet critical manning requirements." 6/

In this case a dichotomy between pay for special recognition and pay to attract additional volunteers is artificial. An increase in hostile pay provides recognition and attracts volunteers at the same time.

The Quadrennial Pay Review criticizes variable hostile fire pay as inadequitable. But their examples are based on soldiers who undergo extreme risks for short periods of time.

"... those involved in fixed battles would have received the higher rate for much longer periods than those suffering greater casualties in more intense yet shorter clashes with the enemy." 7/

If the hostile fire pay resembles the pay schedules described in Figure 1 or Table 1, the soldiers who double the probability of being killed in action but halves the time he risks his life increases his hostile fire pay.

VII.

The probabilities of being killed in action or wounded would help military planning independent of their impact on hostile fire pay. Equipment and men would be shifted from areas with low probabilities of being killed to areas with high probabilities. The impact of different commanders, force structures. strategies, training, on these probabilities would become apparent during a war.

Knowledge of the risks could improve soldier morale given the human tendency to imagine the worst. But even if a more knowledgeable soldier proved more expensive, the <u>social</u> cost would not increase. If the soldier's pay increases, the income of taxpayers decrease by the same amount. The net loss to society is zero.

6/ "Hostile Fire Pay" <u>Report of the 1971 Quadrennial Review of Military Compensation</u>: Office of the Assistant Secretary of Defense (Manpower and Reserve Affairs), December 1971, p.iii.

7/ <u>ibid</u>. p.III.19.

RISK ANALYSIS IN WEAPONS DEVELOPMENT

Mr. Robert C. Banash
US Army Weapons Command, Rock Island, Illinois
Dr. Donald W. Hurta
Defense Systems Management School, Ft. Belvior, Virginia

SUMMARY

The beginning of this report emphasizes the distinction between decision/risk analysis on the one hand and, simply, risk analysis of the other. Briefly, this distinction is that the domain of a risk analysis is only one of the alternative branches of the decision tree in a decision analysis problem. No decision is to be made, the alternative is to be analyzed only in terms of risk.

Next, we suggest guidelines and discuss methodology for performing risk analyses and observe that they are similar, though not identical to those involved in systems analysis studies. The discussion: "When should a risk analyses be performed?" is followed by a discussion of the formation and utilization of a risk analysis task force. This organization consists of a mobile set of "expertise groups" clustered around a program-manager, operations-research core.

Tools necessary to conduct the analysis are categorized as the type which convolve a set of random variables or the type which furnish estimates on the distribution of these variables. Examples are presented to illustrate the techniques required and tools utilized to meet varied risk analysi objectives.

I. INTRODUCTION

The US Army Materiel Command, AMC, has furnished the Commodity Commands with a definition of Decision Risk Analysis (DRA):

> Decision Risk Analysis (DRA) is the discipline of systems analysis which, in a structured manner, provides a meaningful measure of the risks associated with various alternatives, as presented to decision makers. An alternative has risk if there is uncertainty in one or all of the events comprising that alternative.

A set of guidelines on the constituency of a DRA has been provided, (Ref. 5). The key point of the DRA is to aid the program manager in choosing from among a set of alternative courses of action under uncertainty. Risk Analysis (RA) has also been defined by AMC:

> Risk Analysis is any analysis which attempts to quantify or qualify uncertainty.

The suggested guidance is that "Risk Analysis should be an on-going part of every program." A risk analysis may serve to inform the manager of the thickness of ice upon which he walks, but it will not lead him to a decision. It could, however, spark initiation of a DRA. The foundation of the DRA is the ability to assess the risks associated with each alternative; i.e., a risk analysis.

This paper attempts to expand upon the AMC guidance by providing some general remarks on the constituency, tools and application of Risk Analysis and specifically on the policies of implementing Risk Analysis at the US Army Weapons Command. A set of risk analysis guidelines is presented followed by the organization of a RA task force. Tools useful to these analysis are then discussed and excerpts of RA studies are presented to illustrate the application of these concepts.

II. RISK ANALYSIS - GUIDELINES

The assessment of program success is the primary objective of risk analysis. Risk analysis is a process which involves the application of a broad class of qualitative and quantitative techniques for analyzing the uncertainties associated with the realization of cost, time and performance goals of projects. A fourth dimension, risk, has been introduced as a common measure to integrate the three dimensions into a single index of uncertainty and to facilitate trade-offs. One purpose of the risk analysis is to quantitatively estimate this index and determine its variability under changes in basic program assumptions and estimates. Another purpose is the qualitative analysis derived from the network format, if applicable, and the searching questions: What can occur?, What action will be taken as a result of each occurrence?

Another objective of risk analysis is the creation of a quantitative and experimental laboratory to study program success. The general methodology for a risk analysis should be quite similar to the steps involved in systems analysis, systems engineering or industrial dynamics. We propose that the basic risk analysis methodology consist of the following steps:

1. Form a Risk Analysis Task Force
2. Identify objective
3. Specify critical events
4. Develop contingencies for each event
5. Construct program network
6. Collect data
7. Evaluate network
8. State conclusions

It is important to ask ourselves whether or not risk analysis can contribute to acquisition management. A risk analysis can be beneficial by identifying the following areas to program management personnel (Hwang and Meyer 70):

1. potential problem areas
2. consequences of failure

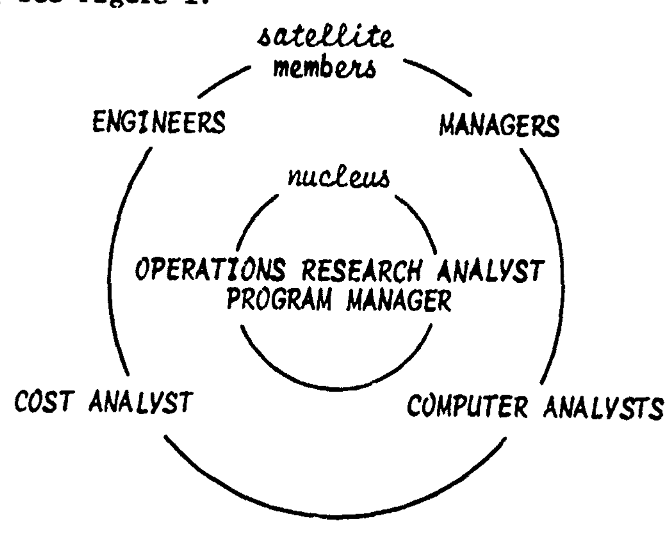
3. low risk program areas
4. requirements versus state-of-the-art trade-offs
5. adequacy of acquisition time
6. sufficiency of appropriations
7. optimum allocation of funds
8. data gaps/recommend studies and concepts
9. sensitive/critical parameters.

One important aspect of risk analysis is to train all program management personnel to become more conscious of system risk. The realization of risk by all levels of personnel would increase the probability of project success and control cost growth, schedule slippages and degradation of performance.

Risk analysis is by nature an iterative process and must be up-dated and validated at regular intervals. It has been proposed that risk analyses be carried out at least during concept formulation, contract definitions and prior to production. However, the greatest benefits will be derived from its use as a framework in which to view current and predicted status. Viewed in this way, risk analyses will be an integral part of the program requiring updating at key milestones of the acquisition cycle and at the discretion of the program manager.

III. RISK ANALYSIS - TASK FORCE

The Risk Analysis Task Force has been structured at WECOM to consist of nucleus and satellite elements. The nucleus has overall study responsibility. It is the continuously identifiable part of the task force. The satellite elements are the managers and technical experts who validate the program network description, contribute direction, and contribute technical expertise as required, See Figure 1.



RISK ANALYSIS TASK FORCE
Figure 1

The nucleus is composed of two key members. One is an operations research analyst, the other a program-manager. The former has the skills required to implement the various risk analysis techniques, while the later understands the objectives and approaches of the development program.

The satellite members enter into the analysis as required. Management enters to assist first in the formulation of program objectives and later in the construction of the program network. Engineering and cost personnel are consulted in the data collection phase to estimate cost, time and performance parameters. Computer analysts may be consulted on the implementation of the selected techniques. Final computer runs may then be made. An analysis of the results by the nucleus members completes the analysis and evaluation phase. Conclusions of the study are presented to management. This may end the study or initiate another.

This type of organization is based on need. Operations research analysts are few in number and are required to participate in several risk analyses simultaneously. The staffs of the program managers are committed to other aspects of the program and are required to participate in the risk analysis while retaining other responsibilities. This organization retains the smallest required number of continuously active participants (2) with knowledge of all aspects of the study, but is capable of expanding to include the active participation of all local experts. The identifiable task forces have been as large as eight active participants, while four is usual. The experts review, critique and contribute to the analysis. The nucleus incorporates these comments allowing the experts to fulfill other committments.

IV. RISK ANALYSIS - TOOLS

The tools of risk analysis may be considered to be divided into two classes: those which assist in the simulation of a suggested program (Construct-Program-Network Phase) and those which aid in extracting basic data (Collection-Data Phase). The former reflects a requirement for summing random variables. The latter extracts estimates on the distributions of these random variables as well as estimates on the end item performance variables.

Systems simulation in its simplest form may be a sum of independent random variables; e.g.,

$$\text{Total Cost} = C_1+C_2+\cdots+C_n ,$$

where C_i = cost resulting from activity "i".

Analytical and numerical techniques exist for determining the "Total Cost" distribution, the technique used depends on the distribution assumptions on the family C_i and the value n. An analytical procedure is used if each C_i is, say assumed normally (or χ^2, or Poisson) distributed. If n is large, the Central Limit Theorem may justify treating total cost

as a sum of normal random variables. Weibull, Beta or Gamma assumptions or mixtures of distributions, may call for numerical techniques. Several computer programs are available. Some are specific with regard to the distribution family, References 9 and 10. Another is quite general with respect to these assumptions, but requires some finesse in handling input to obtain results, (Schlenker, 68). This simulation may result in an end product itself or serve as input data to a more complex simulation.

Summing dependent random variables and variables with infrequently encountered distributions is accomplished by means of MONTE-CARLO simulation. A network describing the program activities is constructed in a form suitable for a canned network analyzer program; e.g., GERT, VERT, MATHNET (See references). The program is "realized" hundreds of times by sampling from the cost (or time) distribution of each activity. Constraints are observed; e.g., activity B may not be initiated before the completion of activity A. These "realizations" yield an empirical distribution of total program cost and time. Examples of these networks are presented under EXAMPLES.

Network simulation techniques are well established. Choosing a network simulation technique, say GERT over VERT, engenders little controversy. (We use VERT.) The choice depends on the peculiarities of the development program, familiarity with technique principles and access to the computer program.

Techniques which elicit a data base for use in the network simulation are, by contrast, controversial. Certain costs, time schedules and performance parameters may be well-known at the time of the risk analysis; e.g., contract price on spare parts purchase, total system weight. Much of the data will be known with less certainty; e.g., time to redesign a replacement part, system reliability. These random variables require characterization by the Risk Analysis Task Force. Tools of the second type are used to elicit "expert opinions" on distributions or parameters of these distributions.

Probability distributions are estimated for cost and time random variable and frequently for performance variables. The family of distributions is either that set provided by the computer technique employed or a family selected by the task force nucleus. (Schlenker, 67 uses a family of Beta distributions.) A distribution shape is selected for each basic activity based on interviews with the "technical experts." Once the distribution shape is selected certain quantiles of the distribution are estimated. The experts may then be asked to provide estimates on, say, the "0", .25, .50, .75, "1.0" quantiles of the distribution, Figure 2. These estimates will usually require adjustment to maintain consistency with the selected shape. Another technique determines the shape of the distribution, from a known family, directly from the quantile estimates, (Cost Analysis Div., WECOM). Alternatively, the expert may be asked to estimate probabilities of meeting or exceeding specified performance values. Time

limitations may call for a qualitative evaluation with risk categorized as high, medium or low in regard to a particular aspect of the program; e.g., performance, as in Figure 3.

The consolidation of these estimates has been achieved by taking the arithmetic mean of the estimates and by attempting to develop a concensus among the experts using the Delphi Technique (Dalkey, 70).

The shakey foundation of the analysis is the basic data. The technical experts may possess little experience in probability practice and so may not be able to translate their "feelings" on chance-of-occurrence-of-an-event into subjective probabilities. It is also presumptuous, on the part of the operations research analysts, to select a probability distribution from the estimation of a few quantiles of the distribution. These problems were recognized and discussed by Grubbs in regard to PERT (Grubbs, 62); this discussion is still pertinent to much current risk analysis practice.

Some of these deficiencies are addressed by performing sensitivity analysis i.e., rerunning the program network-analyzer for various distribution parameters, quantile and probability estimates. This type of analysis is performed in the Evaluate Network Phase. The result is a determination of critical network activities. Observation of a wide fluctuation in the program risk over "reasonable" variations in data assumptions will cause doubt on the meaning of risk as a measure of program failure, but provide the program manager with something to "think about."

IV. EXAMPLES

These examples were chosen to illustrate the variety of technique assemblages employed to address specific risk goals. The first is a two part analysis concerning the XM198, Towed 155 Cannon; the second is an "in process" study concerning the M110E2 Self-Propelled 8" Cannon.

Example 1 (Williams and Banash, 1971)

Objective: Determine current program risks in regard to cost, time and performance constraints.

Techniques:

Simulation: Convolution, of Beta distributions, was performed for total cost and total time distributions using numerical procedures.

Basic Data:

a. Experts selected distribution shapes from family of Beta distributions for cost and time random variables.

b. Performance was divided into five states: 1-unacceptable, 2-marginal, 3-meets requirements, 4-exceeds requirements, 5-meets desired goals. Probability of observing each state was estimated for each performance category.

Figure 2
SAMPLE FORM

Performance Estimates

Component
Definition: θ = Mean Miles/Rounds Between Failures

1. The value of θ estimated for this component is ________.
2. A proposal has been made to replace this component with the component described on the attached sheet.
3. What, in your judgement, is the; a. minimum value of θ _____, b. most likely value of θ _____, b. maximum value of θ _____ for this new component
4. Assume that the value of θ for the new component has been found to be between the values you estimated in 3.a and 3.b. Considering this new information, circle a new estimate of θ on the following scale:

[—I—I—I—I—I—I—I—I—I—]

estimate 3.a estimate 3.b

5. Assume that the value of θ for the new component has been found to be between the values you estimated in 3.b and 3.c. Considering this new information, circle a new estimate of θ on the following scale:

[—I—I—I—I—I—I—I—I—I—]

estimate 3.b estimate 3.c

Figure 3
SAMPLE FORM, ABRIDGED

Assessor ________ Organization ________ Subsystem ________

RISK CATEGORY:

High:
- a. Technical breakthrough required
- b. Design considered to approach state-of-the-art - no prototype
- c. Demonstrated concept in breadboadr - no prototype
- d. . . .

Medium:
- a. Prototype available but functional parameters uncertain
- b. Interfacing problems, but not judged serious
- c. No prototype - but considered well within the state-of-the-art, however requires a cut and try approach
- d. . . .

Low:
- a. Scaling required on a proven design, no problems forseen
- b. Similar component exists - slight modification required
- c. Any competent engineer could build one (requirements are non-rigorous)
- d. . . .

Output:

a. Probability of not meeting time and cost constraints and performance requirements. This was defined as program risk.

b. Risk isograms were constructed which illustrated relationship between time and cost for specified risk levels.

Example 2 (Hurta, 1972)

Objective: Determine technical areas of high risk. Determine solution complexity of problem. (Accomplishment of these objectives would provide guidance on resource allocations.)

Techniques:

Simulation: Not r ired.

Basic Data:

a. A survey was taken to determine existance of uncertainty in areas of technical performance.

Delphi technique was used on these areas to estimate 1) probability of problem occurring, 2) solution complexity according to:

Solution Complexity	Meaning
a	Existing technology
b	Scaled version based on existing technology
c	Limited component test available
d	No lab or component work available
e	No lab or component work and limited theoretical basis.

3) optimistic, pessimistic and most probable solution time.

Example 3

Objective: Assess the risks associated with the M110E2 development program with regard to time, cost and reliability-availability-maintainability, RAM, performance. Identify critical decision points and activities which affect these variables. (Identification of these critical areas was expected to lead to intensive management or consideration of alternative courses of action.)

Approach: The study was separated into two phases.

Phase I ("Rough Cut")

1. establish the scope of the risk assessment,
2. establishing the objectives of the risk assessment,
3. describing the inter-related programs within a network format to define major activities, milestones and decision points,
4. perform a qualitative evaluation.

Phase II

1. quantitative evaluation of the program defined during Phase I,
2. analysis of results,
3. conclusions

Techniques - Phase I

Simulation:

1. Network of program activities constructed, Figure 4.
2. Network entered into STATNET (VERT) Monte-Carlo network analyzer (Moeller, 72).

Basic Data:

1. Minimum, most likely and maximum estimates obtained for cost and time aspects of major program activities. Triangle distribution assumed.

Qualitative and quantitative analysis of the network developed in Phase I led to a determination of critical activities and decision points; e.g., the impact of a timely funding decision would determine entering ET/ST with or without product improved parts. Success of product improvements necessary to program success. This focused attention of ET/ST activities and an expanded network (Phase II).

Techniques - Phase II

Simulation:

1. Network of ET/ST constructed and entered into STATNET program. (Data developed from this served as input to simulation of Phase I.)
2. Monte-Carlo procedure used to convolute cost random variables to obtain activity cost distribution. These data feed both network simulations.

Basic Data:

1. Estimates of quantiles of cost distribution were obtained and averaged over responses.
2. Time quantile estimates obtained, by program analysts.
3. Estimate on event occurrences ("OR" nodes) obtained by averaging estimates of technical experts.

V. CONCLUSIONS

The Risk Analysis is a tool which serves the program manager by attempting to estimate the probability of program success and indicate weakness at the activity level. The task force formed of a small nucleus group which drawing support from management and technical experts on an "as needed" basis provides economical use of personnel resources.

The application of the entire set of guidelines provided in this paper will lead to a comprehensive analysis. However, the objectives, scope and

time limitations of the risk analysis may require application of less than the complete set. These considerations and the complexity of the program under analysis will determine the required tool assemblage.

The basic data forms the usually shakey foundation of these elaborate analysis. The effects of data estimates and assumptions can be mitigated through judicious utilization of sensitivity analysis and interpretation of results.

REFERENCES

1. Grubbs, F. E., "Attempts to Validate Certain PERT Statistics, Journal of the Operations Research Society of America, November 1962.

2. Hurta, D. W., "Determination of Technical Risks for the 155mm Howitzer, XM198," PAA-TN2-72, Systems Analysis Division, US Army Weapons Command, Rock Island, Illinois, March 1972.

3. Hwang, J. D., "A User's Manual to MATHNET," Technical Note SY-TN12-70, Systems Analysis Directorate, HQ, US Army Weapons Command, Rock Island, Illinois, December 1970.

4. Hwang, J. D. and Meyer, A. R., "Analysis of Risk Program for the Refinement of the Materiel Acquisition Process (PROMAP-70)," Technical Note SY-TN11-70, Systems Analysis Directorate, HQ, US Army Weapons Command, Rock Island, Illinois, December 1970.

5. Letter, AMCPA-SA, Subject: Decision Risk Analysis in AMC, 3 June 1972.

6. Moeller, G. L., "VERT - A Tool to Assess Risk," Proceedings of the Eleventh US Army Operations Research Symposium, May 1971.

7. Pritsker, A. A. B. and Happ, W. W., "GERT: Graphical Evaluation and Review Technique - PART I: Fundamentals," Journal of Industrial Engineering, May 1966.

8. Schlenker, G., "An Analytical Estimation of Cost Uncertainty," T.N. 67-3, Weapons Operations Research Office, US Army Weapons Command, Rock Island, Illinois, September 1967.

9. Schlenker, G., "Numerical Methods in Renewal Theory," OR-68-2, Weapons Operations Research Office, US Army Weapons Command, Rock Island, Illinois, February 1968.

10. Uncertainty In Cost Estimating: Methodology," 71-10, Cost Analysis Division, US Army Weapons Command, Rock Island, Illinois, August 1970.

11. Williams, J. H. and Banash, R. C., "Pilot Risk Analysis of the 155mm Howitzer, XM198," Project Manager, 155mm Close Support Artillery Weapon System, US Army Weapons Command, Rock Island, Illinois, May 1970. (Confidential).

A SYSTEM FOR "NEAR REAL TIME CASUALTY ASSESSMENT" IN FIELD EXPERIMENTATION

Walter W. Hollis
Scientific Advisor
USACDEC, Ft Ord, CA.

Introduction

One of the perpetual challenges which faces analysts in whatever field of endeavor is the degree to which their simulations approach "reality." This challenge is particularly perplexing to those engaged in military analysis. On the one hand, many of the customers for military analysis weigh the computer simulations subjectively based on their combat experience and find them wanting in realism. On the other hand when simulations are conducted in the field, there is an equally large group of consumers of the product who criticize the lack of precision because of assumptions made or lack of control introduced in an attempt to increase realism. Finally, field simulations have not been able to generate "pucker factor" realism. This slide depicts the essence of the problem.

Field experimentation has the potential of being a closer simulation of reality than do other techniques available provided one can be in control of the data from the experiment while permitting the participants freedom of tactical action. Real time casualty extraction can add the threat of "kill or be killed" which heightens realism and generates combat-like player reactions. Until recently such freedom of action could not be permiited in two-sided experiments because the requisite means of assuring data control were not at hand. US Army Combat Developments Experimentation Command now has in being a hardware/software system which permits the execution of a two-sided, free maneuver experiment in which the interactions of the two-sides as a consequence of a meaningful assessment of attrition can be examined.

One of the uses of such an experiment is to infer by some predefined measure of effectiveness whether red or blue wins in the long run. Previous experiments concerned with the outcome of one on one duels, experiments with multiple combatants in two-sided duels, and simulations have often resulted in an inferred exchange ratio which has been employed as the indicator of "who won." Exchange ratios based upon previous experiments have been calculated post trial from experimental data confounded by the absence of the player interactions which would occur as a consequence of attrition. For this reason, doubt exists as to their validity. US Army Combat Developments Experimentation Command has recently completed an experiment involving attack helicopters as part of a combined arms team, which was the first experiment in which meaningful real time casualty assessment has been possible. Near real time is a three to seven second delay between the "open fire" event and a notification of duel outcome at the target. This experiment permitted the occurrence of the type of interactions which had previously been missing. Its results together with those of future experiments of the same kind should serve to clarify the honest doubts in the minds of both groups of consumers surrounding past post trial or analytically derived exchange ratios or other similar measures of effectiveness.

As an introduction to the discussion of the hardware/software system required for the extraction of "casualties" during problem play, identification of the data elements required is appropriate. Those data elements drawn directly from the experiment are shown on the next slide. Unambiguous target-firer pairings are, of course, the key requirement. In addition to the data elements shown on the slide, a priori information is required on the probability of kill, given a shot, or for automatic weapons the probability of kill given a burst, for each possible weapon-target pairing. Illustrative of other a priori information required are data on the distribution of times to fire second and subsequent rounds for tank cannon and antitank weapons, for proper control of the rate of fire of such weapons, and data on the basic load of ammunition for each weapon to insure against unrealistically high numbers of engagements by a single weapon.

Instrumentation System

The instrumentation system employed to collect the required

data, process the data and extract the "casualties" consists of a Range Measuring System, Range Timing System, Direct Fire Simulator and a medium scale computer. The latter component both processes data and serves as a control device for the other instrumentation systems.

The Range Measuring System is a omni-directional microwave system operating at 930 MHZ. The system has been designed to provide data from which the position location of cooperative player elements in an experiment may be determined. The system operates on commands generated by the computer and may, in addition to the position location function, perform in the role of an event recorder/ control device through its capacity to transmit to and receive from the player elements simple messages. Position location data are generated based upon the range between transponder equipped player elements (called "B" units in the CDEC terminology), and a minimum of three fixed location interrogator stations, called "A" stations in the CDEC system. The system position location routine polls each "B" unit once a second. At the time of this polling the "B" unit can also notify the interrogator station that there is a message to transmit. In this situation the player "B" unit is again addressed and will transmit its message through the interrogator station to the computer.

The Range Timing System (RTS) provides a common time base for each data recording system. A master timing station transmits IRIG (Inter-Range Instrumentation Group) time throughout the experimentation area.

The Direct Fire Simulator (DFS) is a cooperative laser transmitter/detector system that, when installed on player elements, permits positive identification of firer-target pairings during a trial. The transmitter is a low power gallium arsenide laser operating at 9050 A^{o}. The beam width is five milli-radians. The output energy of this transmitter is such that no eye hazard exists which restricts its use in a two-sided problem. Each laser transmitter emits a uniquely coded pulse which can be detected and decoded at the target. The laser sensors are silicon photodetector diodes. The DFS interfaces with the data link of RMS so that firer and target data can be transmitted to the computer. It is worthwhile to note that the name DFS is really a misnomer. The DFS is a communication link for identifying firer and target and in no way simulates the characteristics of any weapon.

The computer employed is the General Electric Model 605 which has a 64,000 word core memory. The normal compliment of peripheral equipment such as line printer are also available. The computer serves to control the instrumentation system in accordance with its program and to perform the calculations requisite to the "near real time" casualty assessment routine. The interconnection of these systems is shown schematically in the next slide.

Casualty Assessment Routine

As noted earlier, the casualty assessment routine is based, in part, on data directly drawn from the experiment and in part on data drawn from sources external to the experiment. The casualty assessment routine operates typically as follows:

A target is taken under fire by an attacker system employing its organic sighting system to which the laser transmitter of the Direct Fire Simulator System has been boresighted. Upon completion of the weapon laying process the gunner activates his weapon firing system which triggers the laser transmitter. On firing of the laser, a coded signal is sent to the control computer through the RMS system identifying the laser which has fired. The firing event time is recorded based upon the IRIG timing system.

Assuming the laser beam illuminates the laser detectors mounted on a target, the fact of the receipt of the laser signal, the identity of the firing laser and the identity of the target will be transmitted to the central computer via the RMS system. The crew of the target vehicle is also provided a visual indication that the vehicle is under "fire."

Since, in addition to the event data, the RMS also provides data from which the position of both the target and the firer may be derived, the computer has, at this step, the identity of the firing weapon, the identity of the target and data from which to compute the range between the target and firer.

The range between target and firer is computed, a look-up table of probability of kill given a shot for the identified target-firer pairing is entered and a specific P_k selected. A random

number is then generated, compared to the P_k number and the kill/no kill decision made.

In the event of a kill, the RMS sends back to the target a signal which triggers an audio-signal indicating a kill to the target crew. This signal also disables the laser transmitter on the target. A manually triggered smoke grenade is fired on receipt of a kill signal to indicate to other players that a casualty has occurred.

The basic load of ammunition for each weapon is stored in computer memory. At each firing the firer's ammunition supply is decremented by one round. When the basic load is expended that player is notified through the RMS data link of his out of ammunition condition and his laser transmitter is disabled.

For non-automatic fire weapons such as tank cannon and command to line of sight missiles such as TOW, the time to fire second and subsequent rounds is determined by a controller with the system. This controller draws a number from distribution of data on times to fire second and subsequent rounds. By this technique a realistic control is exercised over rate of fire.

Although the basic sequence of operations for the casualty assessment routine is the same for all weapons, there are specific variations or additions which are dependent upon weapon type. Some of the more significant are described below.

For command to line of sight guided missiles, the firing simulation ground-to-ground and air-to-ground requires that the firer be in laser contact with the target in one of the last three seconds of flight prior to impact. This requirement is to insure the effect of flight time is not washed out of the engagement and as a reasonable means of insuring that the gunners track the target and are properly exposed to counterfire during time of flight. After receipt of a signal at the computer that launch has occurred, and if a target has been identified, a countdown starts based on the time of flight of the missile to the indicated range to the target. Beginning at time of flight seconds minus three and continuing to time of impact, a check is made for laser communication with the target. If the communication check is positive, the shot is processed for a casualty outcome. If not positive, the missile is a "no hit." Ammo supply is decremented and the firer is notified

of missile impact. In the event of a firing without a target identification, the firer will be notified at the expiration of the missile maximum time of flight that impact has occurred and his ammunition supply is decremented.

The probability of kill given a burst for the predicted fire antiaircraft weapons is computed by means of a Salvo fire equation. Data drawn from the experiment for this computation, in addition to the range between the target and firer, are the velocity of the aircraft target and which of seven different aircraft aspects is presented to the firer. Velocity information is extracted from the time based position location data. The seven target aspects are derived from three independent groups of laser sensors mounted orthogonally on the aircraft. The specific target aspect is determined based upon whether one, two or all three groups of sensors are illuminated. This information is transmitted to the computer via the RMS data link.

The laser beam divergence (5 mils) of the DFS necessitates the generation of certain rules of engagement. These are:

If a single target is illuminated simultaneously by more than one firer only one firer will be processed.

If more than one target is illuminated by a single firer, the closest target will be chosen by the computer system logic. This rule was selected since it appeared not to favor either side: Other rules can, of course, be played. For example, the selection of a priority target such as an air defense weapon in the case of an attack helicopter.

Conclusion

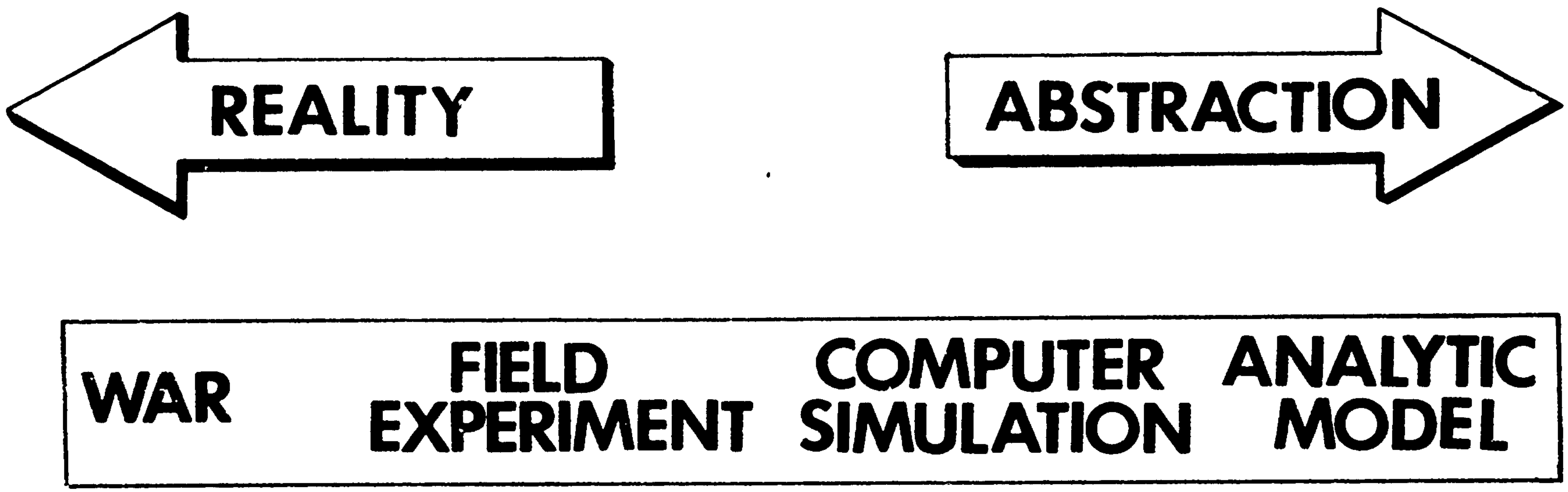
The next slide shows a portion of the result of a typical experiment trial displayed on a time line. Each engagement is identified by a vertical line extending between the firer and his target. The engagement range is given in meters. The engagement outcome, survive, kill, out of range or other appropriate outcome is given by the symbol at the target end of the line. The symbols are defined in the legend at the bottom of the chart. The time scale is one minute between numbered divisions. This type

of data display is one of the two key data displays required for analysis of experiment outcome. This display is critical, since it shows on a trial by trial basis the degree to which the data on each exchange of fire are unambiguous. The other key data display is the computer print out which shows the P_k assigned to each exchange of fire.

The primary purpose of casualty extraction is to permit "realistic" interaction to occur between the players within and between the opposing forces. The observed number of blue and red casualties by trial or series of trials is not used per se to formulate an estimate of the loss exchange ratio. An expected value for this ratio is computed based upon appropriate summation of the $P_{k's}$ overall engagements.

The instrumentation used in the initial experiments was essentially prototype and, because of vehicle vibration and other electronic interference effects, there were a number of engagements in which there were no corresponding DFS laser hits on targets. However, the instrumentation provided so much data associated with each event that post trial analysis was able to verify 96% of all events. The preliminary results of this analysis, 10 trials out of 43, showed that approximately 70% of all casualties were extracted in real time. This post trial analysis provided extensive data, voice, film, ground track, and controller observations on more than 3,000 events in simulated battle.

Given the ability to perform this kind of combat simulation and collect this kind of data with real people on real terrain it appears that those engaged in the digital combat simulation discussed earlier could begin to build bridges between the two types of simulation that would increase considerably the utility of both types of simulation. Some work of this type is ongoing already in connection with studies of the attack helicopter materiel need by the Army. It appears that the type of field experimentation described could be used to verify or to refine models in common use today. Since field experimentation is expensive, the construction of relationships between the two types of simulation requires careful thought and analysis, so that such field experiments as may be identified for execution are addressing the most relevant questions. In this regard a statement of Mr. D. C. Hardison, Combat Developments Command Scientific Advisor, bears repeating, "What is it I need to know, that I don't know, that I can learn through experiment?"


REALITY
ABSTRACTION
WAR
FIELD EXPERIMENT
COMPUTER SIMULATION
ANALYTIC MODEL

1

OBSERVED PERFORMANCE DATA

Time of TRIAL INITIATION

Time of INITIAL DETECTION by DEFENDER/AGGRESSOR

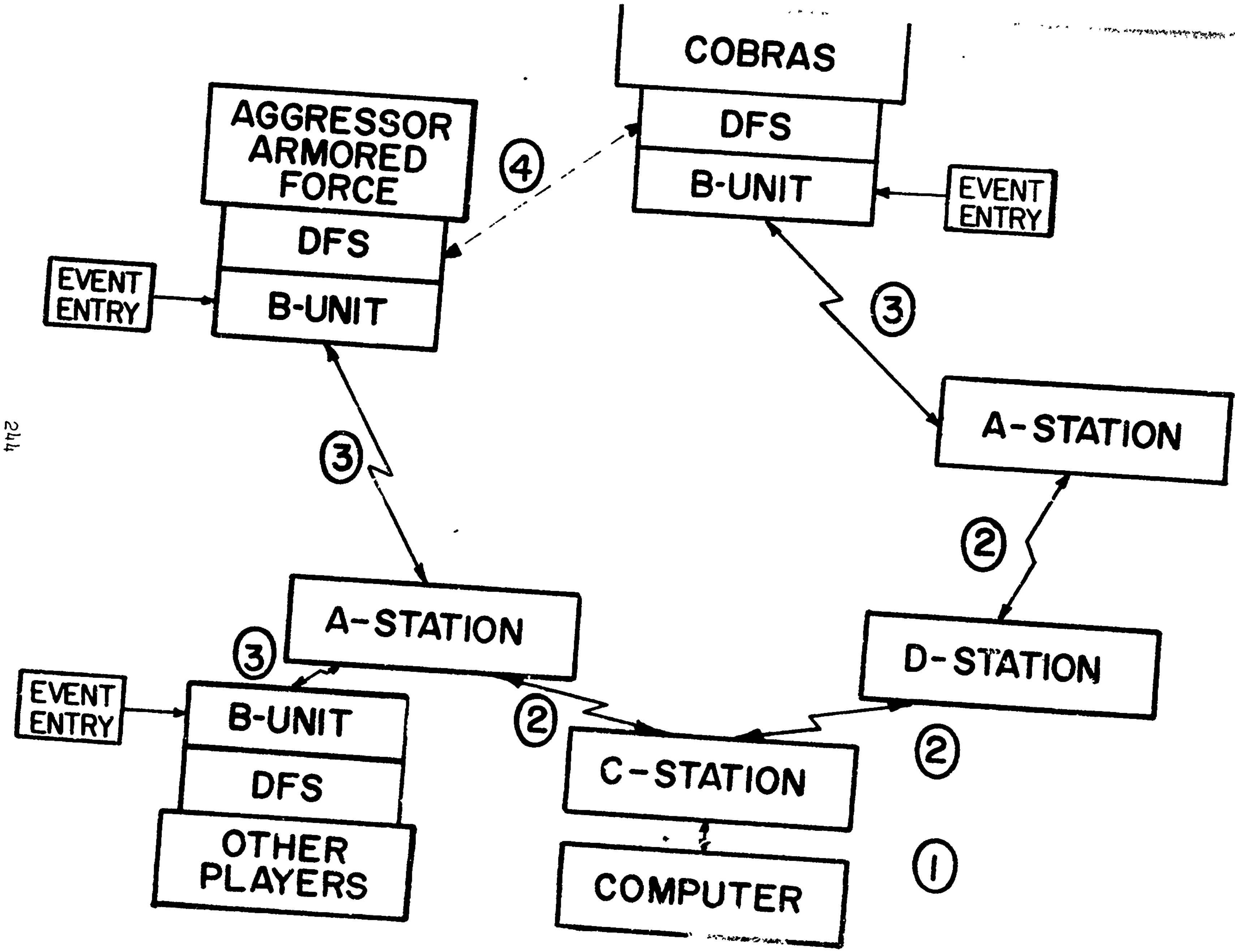
Time of each WEAPON FIRING

IDENTIFICATION of firing WEAPON and TARGET in each ENGAGEMENT

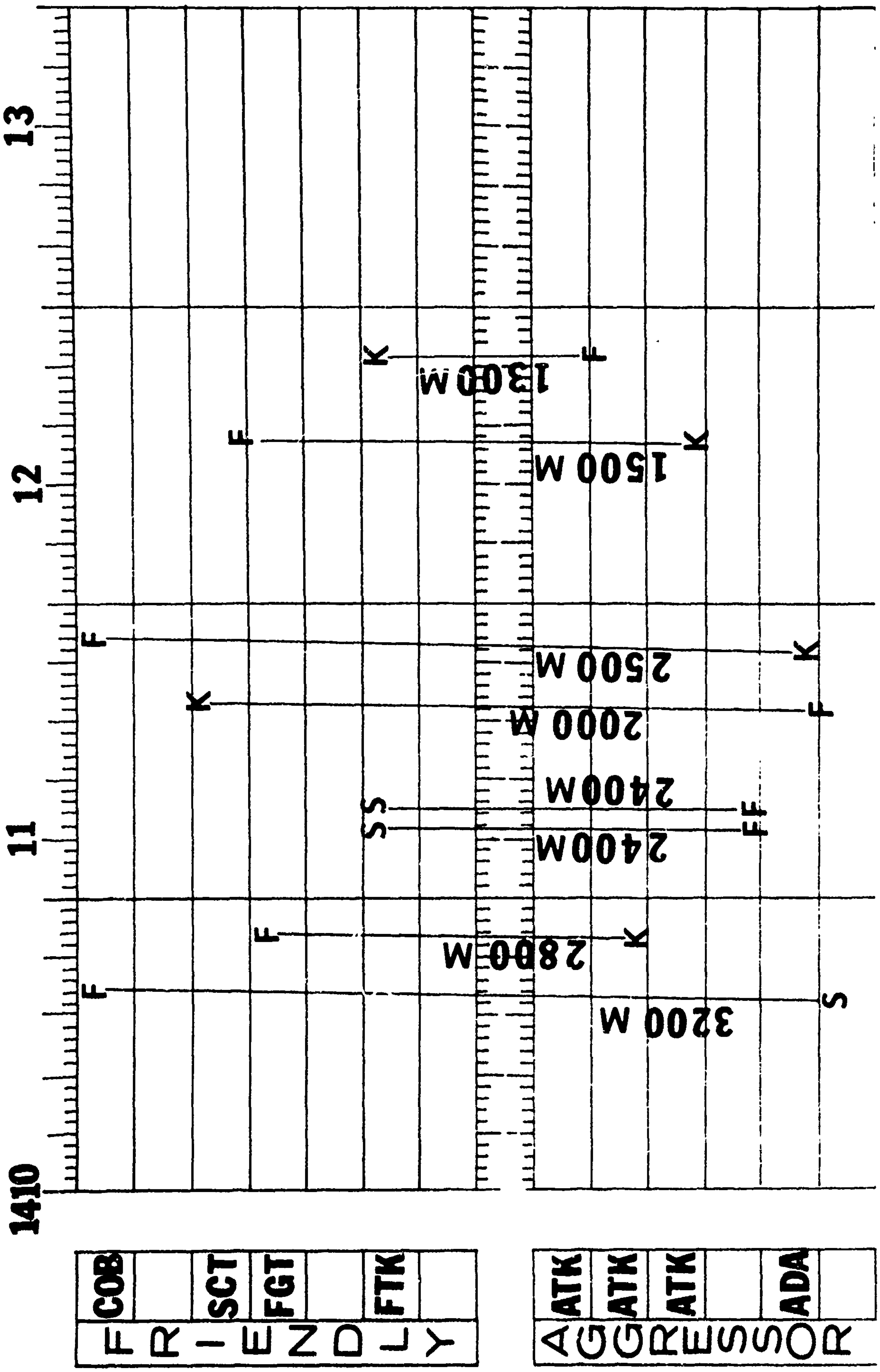
END of FIRE TIME for each ENGAGEMENT

TARGET ASPECT PRESENTED for each ENGAGEMENT

RANGE of each ENGAGEMENT


COBRAS
DFS
B-UNIT
EVENT ENTRY
AGGRESSOR ARMORED FORCE
DFS
B-UNIT
EVENT ENTRY
4
3
3
A-STATION
2
D-STATION
2
A-STATION
3
EVENT ENTRY
B-UNIT
DFS
OTHER PLAYERS
2
C-STATION
COMPUTER
1

SLIDE 3



13
12
11
1410
1300M
1500 M
2500 M
2000 M
2400M
2400M
2800 M
3200 M
COB
SCT
FGT
FTK
FRIENDLY
ATK
ATK
ATK
ADA
AGGRESSOR

INTRODUCTION OF BANQUET SPEAKER

by

BRIGADIER GENERAL CHARLES D. DANIEL, JR.
Director of Army Research

In taking a quick look at the background of the past speakers of the Operations Research banquet I found that, of course, the backgrounds come from a broad range of occupations and the people who spoke, spoke on diverse topics. While these topics were not always OR oriented they all had one thing in common and that is a real strong interest on the part of the audience, in particular, and the part of our nation as a whole. As an example, our speaker in 1962 was Dr. Paul Gross of Duke University and he spoke on the subject of "The Challenge to Science in the Second Half of the Twentieth Century." In 1963, Dr. Hugh J. Miser of The Mitre Corporation spoke on the subject of "Operations Research". In 1965, Dr. Wernher von Braun, spoke on the subject of "Keeping a Date in Space". In 1971, we had Mr. William H. Megonnell of the Environmental Protection Agency who spoke on "Federal Air Pollution Control Program: An Evolving Blueprint." Tonight our speaker is Brigadier General Robert Gard who is the Director of Discipline and Drug Policies for the Department of the Army and he will speak on a subject which has the same sense of urgency and the same high level of interest to us, either as members of the military, as parents, or as members of a great nation that finds itself in a very precarious position. The subject that he will speak on is "Drug and Alcohol Abuse-The Army Drug Program."

Before I let General Gard take over the rostrum, I'd like to give you an idea of his background so you will appreciate the depth of experience that he brings to this audience. General Gard is a graduate of the United States Military Academy, the Command and General Staff College, and the National War College. He received a Master's degree from Harvard University in 1957, and returned to the Military Academy to join the faculty in the Social Sciences Department. At the invitation of the Dean of the Graduate School of Public Administration, General Gard returned to Harvard University for the academic year 1960-61 and completed his Ph.D. in political economy and government. General Gard is also obviously a soldier and he's in the best branch of the Army, the Field Artillery. I always have to get in a plug for the artillery. If we don't speak for ourselves nobody else will. His troop assignments include duty as a battery commander in the Korean War, and a 105 mm. howitzer battalion commander in Europe. He was the division artillery commander of the Ninth Infantry Division in Vietnam from August 1968 to May of 1969. He's served in several high level staff positions including the Office of the Secretary of Defense where he was a Staff Officer in the Office of the Assistant Secretary of Defense for International Security Affairs, later Special Assistant to the Assistant Secretary, and finally, Military Assistant to the Secretary of Defense. It's with a great deal of pleasure that I introduce Brigadier General Robert Gard.

Preceding page blank

DRUG AND ALCOHOL ABUSE-THE ARMY DRUG PROGRAM

by

BRIGADIER GENERAL ROBERT G. GARD, JR.
Director of Discipline and Drug Policies
Office of the Deputy Chief of Staff for Personnel

Thank you very much Charlie, and my fellow drug abusers and those few of you who have never had a hangover. I was a little uneasy about coming down here to address this group because I know of your high competence in things I don't understand, but I was much comforted this evening to run into a lot of old friends who were down here under the same false colors I am; they don't know anything about what you're doing either so I feel much more at home. The letter that was sent to me to invite me to meet with you here this evening mentioned that it wasn't necessary that I be a member of the Operations Research community and I was grateful for that. I want to warn you, however, that the people who planned this meeting have a more insidious purpose in mind because my selection for this honor is related to the theme of this conference. Permitting me to speak here tonight is a case study in "risk analysis," and just to let you know that the administration understands this, the fact sheet that I received, and I think you did too, singled out this meeting this evening in that it felt it necessary in view of the speaker and the topic to caution-and I now quote-"the banquet is considered a part of the symposium and attendance of all participants in encouraged." I think I understand what they're saying. I do want you to know that I appreciate your forbearance in coming out this evening to listen to this particular subject area because it's one that has certainly occupied my time over the last several months and one that I think affects us all. We've really had a serious drug problem for some time in our society and in the Army and I refer, of course, to my favorite drug, one that some of us abused this evening to the extent that, as our master of ceremonies mentioned. they ran out of it. So if the shoe fits, put it on.

This society tolerates 38,000 alcohol related deaths a year. There are, depending on your definition of alcoholism, upwards of 5 million alcohol addicts, we call them alcoholics in this society. Many hundreds of thousands to the point of personal and family tragedy. There are thousands more, indeed millions more, of problem drinkers whose use of alcohol creates difficulties for them and surely alcohol is the society's most serious drug of abuse. Then, of course, as all of us know because we are in a drug-related society, the housewives who take downers to help them sleep at night and maybe uppers to help them go in the morning, they're not junkies but they've got a drug problem too. So what's the big concern? Our society is accustomed to paying this high price for drug abuse. To produce a perception of a social problem, I would contend that the dangers must be generally unfamiliar; they must be relatively widespread; drug use in a manner that deviates from the accepted social and medical patterns within a given culture; and most important they must produce significant antisocial behavior that the society finds intolerant.

Preceding page blank

Now, if you go back a bit to look at the evolution of our recent drug problem, you know you go back to the early 1960's up toward the mid 1960's when LSD, Timothy Leary, and the flower children were prevalent. These were dangerous and unfamiliar drugs. There was some loss of life, but basically the flower children were quiet and peaceful and introspective and society didn't worry very much. Then if you take Haight-Ashbury as kind of the community that leads in this field, although I was told tonight that this particular community has been somewhat progressive in this regard as well, you move into the use of "speed" along about 1967. The speed freaks were paranoid, aggressive and violent. You begin to get some antisocial behavior that concerned others. The dope heads, people who used the downers, then transitioned into heroin and you can pick your dates-1968, 1969. The number of reported deaths began to go up, a thousand a year in New York alone. The spread became rapid throughout society, it left the ghettos and reached the suburbs and people began to be concerned. Concerned because the degree of unfamiliar antisocial behavior became sufficiently widespread to get our attention. I refer, of course, to the significant increase in crime that was related to people maintaining a habit on heroin and the public concern that then grew up was reflected in the Congress, evidenced by the reports of Congressmen Murphy and Steele who went to Vietnam and reported drug addiction among our GI's at 20 to 30%. Then finally about a year ago, the President declared a national counter-offensive against drug abuse. He called drug abuse "public enemy number one." So we, in the military, who had tended, I think, to sweep this problem under the rug somewhat, much as we had the alcohol problem, found ourselves confronted with a situation, not a theory, and the priority, of course, was on Vietnam because it was there that this very potent heroin was readily available. It mushroomed in that theatre and about this time period and we were told to identify, immediately, those who were dependent on heroin, treat them and not return them to streets already with an intolerable crime rate. That's the priority we received. We began screening soldiers leaving Vietnam the day after the President's announcement and this was really the beginning of anything you could call a comprehensive Department of the Army attempt to meet this very perplexing drug abuse problem. There were some previous efforts. In the fall of 1970, we published an Army Regulation which directed education in drug abuse and permitted commanders to offer amnesty to those who might seek treatment, but basically we approached this problem from a law enforcement standpoint. We were forced, when the Executive became energized to begin to do something about this, to start a program not only before we could provide trained personnel and resources to the field, but even really before we could plan the program.

With that introduction I thought what I might do is sketch for you how we met this particular challenge in the Department of the Army; sketch for you a brief outline of concept of our program and then give you the opportunity to ask questions of me in those particular areas of the program that you find of particular interest or concern. As I mentioned a moment ago, we began in Vietnam, that's where it all started, and you might enjoy this because this is not unique to me and you'll recognize things that have happened to many of you. I was having a marvelous time in New York City at the Council on Foreign Relations, a very relaxed kind of assignment and I had seen an article in the New York Times, toward the end of May, and it showed a picture

of the Joint Chiefs of Staff and the President, who was looking very somber. The President was talking to the Chiefs about the forthcoming drug abuse counteroffensive in which the military was to participate, and I told a couple of friends of mine at the council; I pointed to the picture and said, "Some poor soul is going to be jumping in about two weeks." Two weeks later I got a phone call and someone said, "Sir, we have an assignment for you." I said, "Great, I've been trying to get one out of you for six months. What am I going to do?" He said, "Well, don't delay, we want you down here in two days. You're going to run the drug program for the Army." I said, "Man, you've got to be putting me on." I went down there; I had heard of heroin, but I didn't know what it was, and two days later I was negotiating with the White House on our terms of reference. This thing really moved fast. Nine days after my arrival the President made his announcement and the tenth day we were screening people in Vietnam. After we screened these people and found out who they were we had to medically evacuate them to the States. The question then arises, even to a guy who knew as little about it as I did, what are we going to do with them when they bring them back? We didn't have any drug treatment centers. I had negotiated with the Veteran's Administration about how we would phase people into their facilities and they announced to me they only had five drug treatment centers. I said, "That's very interesting, we don't have any." What we did was go out to the 34 largest Army hospitals and said, "Guess what? You're confronted with a situation, not a theory." We picked these hospitals because each of them had a mental hygiene clinic, which meant they had a psychiatrist and a social worker and some technicians to assist in the program so when the men came back they would have somebody who could assist them. This whole program was as new to our medical personnel as it was to the rest of the Army. We decided right away that we were going to decentralize this program. We considered other alternatives but we decided on decentralization for two reasons. First, because the prognosis for rehabilitation seemed to be more promising, but in my view more important, we made this choice because the drug problem is epidemic in the military communities just as it is in the civilian communities. There was really no practical way that you could allow a commander the luxury of saying, "I have a man with a drug problem, ship him off someplace and cure him and then send him back to me." That would have been an understandable reaction for a commander, particularly in this time period, faced with all the difficulties of troop command, but we decentralized and we did so very deliberately. Now the soldier would be medically evacuated and this was the beginning of the program. If he was due to get out of the service, we sent him to the hospital nearest his home; if he was due to stay in the service we sent him to the hospital nearest his next duty station. We than tried to stabilize him in the hospital, as I'll cover a bit later, and tried to get him back to work in the unit as quickly as we could.

You may be interested in whether Murphy and Steele were right. Were there 20 to 30% addicts in Vietnam? Actually, it's really hard to say and I'm not confident about the figures early in the program and I'll tell you why; because we sent these test machines over to Vietnam and we had them in place and began testing a day after, but initially we didn't have the lab procedures down very well. I thought like when you collect a urine sample you throw a pill in it and hold it up to the light. "That's green, man it's

go, he's okay." If it's red, you've got drug problems. It's very complicated and we didn't have this shaken down by any means, initially. I say all this to let you know when I indicate to you that we're running about 5% positives for the people coming home you can put it in context because it took us a while to shake the labs down. We, of course, put control samples in the labs to check them out and found out that they weren't nearly as accurate as we had hoped. And, of course, the individual on his way home knew when he was going to get tested so if he could abstain from using heroin for a matter of a few days he could slide under the screen and come home or the freedom list. It was only a few weeks later that we put an unannounced testing policy in Vietnam where we went out into the units and by that time, of course, the word was around that we had this particular program. We tried to do something from the standpoint of education so we weren't able really to measure what the prevalence was in Vietnam because by the time we began the unit screen, even though individual units would vary all the way from say 0 to 25% in the cases of some companies, the average again was slightly over 5%. Where have those figures gone since then? Well, by February, on the people coming home we dropped down from about 5% to 2 1/2%, and I'm pleased to tell you in March it was under 2%. In the case of the unit test, it stayed around 5%, but you must understand first, that the labs got better, and secondly, those units that tested with a high prevalence, initially, got tested every two weeks. This biases the figures upward. You may have seen some stuff in the press that said they haven't gotten any better because it's still 5%; but in fact we have, because if you take out the units that we repeat the test on and just kind of normalize the rest of them, it's now running down around 3%. This caused the Assistant Secretary of Defense of Health and Environment who visited there to say, and I was very grateful to him, "The green machine has done something right." You know, if he had said, "The Army program had positive results," nobody would have printed it. He made every paper with the "green machine" doing something right. We, I think, have done something about at least turning that epidemic around in Vietnam. Meanwhile back in the States, of course, here come these people and within six weeks after the President's announcement we'd contracted with civilian laboratories and had begun a program of special event testing in the States. People coming into the Army get tested, people on their way overseas get tested, people getting out of the Army get tested, and again you may be interested in kind of levels of magnitude that we found. Let me caution you once more that these labs turned out not to be as reliable as we'd hoped they'd be. They were servicing drug programs where they had smaller numbers of samples with a high drug content and that gives them one kind of testing problem and one kind of management problem, and, if you'll pardon the pun, when we flooded the labs with the urine samples and they had to test large numbers of them, it became a little different game. They weren't really equipped to handle samples of this magnitude and as we slid control samples in, and so forth, we found out they were flunking the course. So, at least, initially, we were getting an understatement-to what extent we're really not certain-of the problems, but with some understatement. I might add, by the way, that we have since written those contracts and put a financial incentive in. If we slide control samples into a batch and they flunk the course in the control samples, then we get paid for the whole batch. This has given them incentive to improve their management techniques in the lab. We haven't solved that

yet so we're putting contact teams out in each of these three labs that we're using in the States in hopes that we can improve their performance. Basically, it was running about 1.5% of people coming in, that's lab positives. Some of these guys may have been taking GI gin or paregoric for other kinds of ailments and we weren't able really to go through the statistical battle of trying to correlate the lab positives with each individual because that hassles the field even greater than we're hassling them now; so these are just lab results. But there are about 1.5% of guys coming in and about 1.5% of guys going overseas; no increase yet, but about .4 percent increase when people are getting ready to get out of the Army. We also test those undergoing rehabilitation back in the States and, of course, the percent positive is much higher, but not the 90% that was predicted by some. It's been ranging somewhere between 10 and 15% which is really so phenomenal that I just want to leave that now and I'll mention it again later.

Ten weeks after the President's announcement, we were testing all overseas returnees, the same program we had in Vietnam for the people coming home. Two months after that we had begun unannounced tests, treatment and rehabilitation efforts on station worldwide. I don't mean to stand here and tell you that we have a highly professional program on every one of our installations all over the world. We certainly do not. It's like anything else that you do on a decentralized basis. Some people respond very well and very quickly and others are slower. We felt it was better to push hard to get this going, because we'd ignored it too long, than to delay until we had optimum kinds of situations in which we could implement the program. That's the quick overview of how we responded and we leaned toward perhaps going faster than a lot of people felt we should, but we felt it was worthwhile trying to get going.

Now just let me outline to you how we conceive of this program, and I think it's more simple to do it functionally. We start off with prevention, identification, treatment, rehabilitation, evaluation a d research.

First, prevention. Law enforcement, of course, plays a role here, basically to suppress the supply, to catch the pusher, to punish in appropriate cases, to deter use by those whom these procedures may influence; but more importantly through education. We try through our education program to reach all the different target audiences within a military community and it's our belief, although by no means certain of this, that probably you have to do different things with these different target groups. For the young soldier we start with the credibility gap, not only because of our age but also because at the outset of the marijuana problem we exaggerated to such an extent he doesn't believe us anymore. We have an institutional obligation to make certain that no soldier begins or continues the use of any drug out of ignorance. We owe him at least what we know and what we don't know about this problem so he can make a rational decision. For those soldiers for whom drugs are not the problem, but rather the solution to other problems, we need to make certain that he's aware of alternative means to meet his difficulties. Now with the senior leaders and supervisors you not only have to worry

about teaching the technical aspects of drugs, the drug scene, the drug culture, causes of drug abuse, but we have to establish, very clearly, the requirement to create an environment in which drug abuse is less likely to occur and when an individual has a problem he's got the maximum opportunity to solve it. That involves a whole set of things that some people have grouped under rubric called the Modern Volunteer Army, which I won't go into, but hopefully we'll attack this from the positive standpoint as well as just treating the symptoms of what are really manifestations of an inherent social problem. Now for the young leader I suspect you may have a different problem because, after all, he relates to the young soldier yet at the same time he's in a position of authority and has the most difficult leadership problem. Then, of course, we have our Department of the Army civilians and we have our dependents, to include school children. I've dealt somewhat on the education part of this more than I will the rest of it because we do feel this is really a keystone to the problem. Both for prevention, in so far as that can be done and also to set the proper stage for rehabilitative effort.

Secondly, identification. We do this, as I've already mentioned, through a mandatory urinalysis program where we can detect the presence of amphetamines, barbiturates and opium in the urine samples. Of course, there are other means, the unit commander's observation, and this applies particularly to abuse of alcohol-that's one we can identify fairly readily-but to the other drugs as well. There are certain characteristics with drug user's abuse that are distinguishable and also apprehensive when your men are picked up by the police. But preferably the individual volunteers for assistance when you recognize that he has the problem. We have a policy that we now call "exemption", that we used to call "amnesty". We changed the name because amnesty seemed to suggest kind of total forgiveness for everything. So we changed to exemption not because that helps a lot, but because that's the term that the Office of Secretary of Defense used and it does at least connote that we're exempting the man from prosecution under the Uniform Code of Military Justice or if his inability to kick the habit requires his release from active duty prior to his expiration of term of service, that discharge will not be under less than honorable conditions, solely on the basis of his individual use of drugs or possession incident to that use. We hope also, of course, that we can overcome some of the inherent disincentives in this exemption program, which I won't go into now but will be glad to discuss if anyone cares to raise it later.

Prevention, identification and by whatever means we identify the man every soldier with this kind of problem deserves detoxification, if required, and whatever initial medical treatment may be necessary. I would just mention here that our concept is to treat people on an in-patient basis for the minimum essential period and this leads me into rehabilitation and it's really an extension of the initial parts of treatment.

Our concept here is to try to return the soldier whose term of service has not expired to full and effective duty as soon as possible and for the individual who is due for a discharge to insure his continuity of treatment, and I'll explain how we do that in a moment. Our concept is to place the man in a unit where he can utilize his skill, can function as normally as

possible as quickly as he can. We use so called "halfway houses" for transitional kind of treatment. Perhaps the individual who has been in the hospital is not really ready yet to face the unit environment, totally, and needs some brief period of time in a halfway house. He works out but lives in-again for a brief period-before he goes back to his unit. We use both the halfway house and the rap center which could be a part of the halfway house or a separate installation. So the man receives out-patient counseling of some kind or another; later on, group therapy, whatever may fit his individual case. We're trying very deliberately to allow local initiative in specific modalities of rehabilitation, both because we really don't know what works there-I think that's one thing that's very clear in the whole drug problem-but also because we must capitalize on what local talent is available and have them do the things they know how to do best.

Now let me move into evaluation. There is, as I suggested, no conventional wisdom nor generally accepted solution to this particular problem so we developed a rather extensive evaluation program beyond the usual bureaucratic requirement for management information for a high priority program, and those of you out in the field who have had to render reports have suffered under this. It's an extensive data collection effort. We have Department of the Army teams that we're sending out to the fields basically to gather data to try to give us some indication of what's happened. This leads me into the research program and just like education I want to spend a bit more time on that because I think that of particular interest to this group.

In part, the research program supports the evaluated effort, but I found out in dealing with the people in the research community that there is not the extent of interest in evaluation as a part of research that I'd hoped there would be. It's somehow a foster child in a way. It's not considered really legitimate research and that's been one of our problems. There's a medical R&D program of about three million dollars this fiscal year. It's got three principal parts to it. The first one is an evaluation of the efficacy of the medical aspects of the program, attempting to look at the effectiveness of detoxification-we have a rather extended program in Vietnam of really studying the detoxification process among the heroin abusers-and also to look at the various treatment and rehabilitation programs. The Surgeon General has teams going out to contact the field to try to get a little better idea of what modalities are being used and attempt to determine those that seem to be working best. This portion of the program runs about one million dollars this fiscal year and it will go up to 1.6 million dollars next fiscal year. They are also trying to establish improved laboratory techniques and the development of new technology to detect the presence of drugs, biochemically. This is running in at 1.2 million dollars this fiscal year and will go to 3.8 in FY 73.

The third and kind of general category-I'm grouping all this together-is called prevalence, incidence and medical complications of drug abuse. We're trying to look at personal, social and environmental factors. They had such specific projects as an epidemiological study on the distribution of drugs; the competition between the informal groups of drug abusers and the bureaucracy itself; profiles of soldiers who are involved in drugs, particularly in Vietnam;

and effects of alcohol and drugs on the biological fetus position, metabolism and effects on the brain and behavior and this, by the way, includes marijuana. This is only running $850,000 this fiscal year, but it jumps to 1.9 million next fiscal year. The whole medical R&D program which has been at about three million dollars this year more than doubles to over 7 million dollars and if we can believe the long range forecasts it's going to stay at that level through the next five years. We also have a behavioral science research program under the aegis of OCRD, our host this evening. We are attempting to assess the educational program. We've put a lot of capital in trying to develop programs to reach those educational objectives that I mentioned earlier and we found out very early on that there simply was a critical shortage throughout the Army of people sufficiently qualified to communicate creditably in the whole area of drug abuse. We joined forces with Yale University, which had a military faculty combined with the civilian faculty, and put on, under the sponsorship of OCRD, a two week kind of experimental course of total immersion for the field. We went out to the field and said, "Look, we're not going to train experts and put them on the road to spread the gospel. You send us four-man teams." We picked four men because we sat down and thought what we'd like to have as a commander. A commander would probably like to have more, but we couldn't spare that number so we came up with four. We said, "Send us four-man teams that you want to be the educational cadre for your command and we'll give them the best we've got for two weeks and send them back to you to operate under your command or your education program." We trained about two hundred people this way, military and civilian, not just for the major commands and the larger subordinate commands but we also ran a cycle for the instructors in the service schools because we recognized that this is a kind of long-term, broad-range educational effort and we felt we'd make a lot of money by having people trained in this way in the service schools. But what we want to find out now by our research effort or evaluation effort is what really works when you get out in the field. So we designed a program, with the support of OCRD, to run an assessment of what's really happened in the field. That is what I'd like to see them come up with. Again, as a layman here I don't know if this is researchable and I've spent a lot of hours trying to find out. I'd like when they come back at the end of a period of time, nine months to a year, for them to tell me, "Look, if you're trying to talk to senior noncommissioned officers, technique A works better than B, B works better than C, and technique D bombed", and do this for each of these different target groups. Don't try to put a percent of effectiveness on any of them. I don't know if you can build a typology if you can generalize to that extent. I don't know to what extent you make caeteris paribus in this particular area, but I just feel very strongly that we have to try to find out so we can go back to the field after a period of time and say "Look, in your drug education, we find that this is the kind of technique that seems to get to these various target audiences." That's a big program. It's only funded at three man years-three professional man years-this year, but it goes up to seven next year. No contracts have been let as yet; I'm very anxious to get this one underway.

The second major category, the drug abuser's perception of voluntary disengagement. This is something that the research types thought up and sounds to me like a fascinating idea to try to find out what makes the guy quit using drugs on his own, if we can determine that. That's going to be funded at three professional man years.

And the third general category is organizational factors significantly associated with drug abuse. What they're going to try to do is look at units with high prevalence use and low prevalence and try to find out what the differences are. I'm not as excited about that, although I'm perfectly willing to have the results because I think I know most of that answer. I would bet more on the answer to that one than I would the other two. But that's being done three man years this year and three man years next year. If you add up the FY 72 program, it's under 1/2 million dollars. FY 73 they're going to expand with a new program which is related to the one I just discussed. That is, to try to look at the role of company level leaders in preventing drug abuse. There are no papers on this yet but it's programmed for two professional man years for FY 73 which, when you add up what's going to be carried on in FY 73, it's slightly over 1/2 million. That's the level of research efforts that we're doing.

Let me just conclude by saying that we're really into an area about which this society knows very little. The task in many ways is often difficult, however, I think the opportunities are very great. People like Senator Hughes, the White House office, indeed even the President are looking to the Department of Defense, and most specifically the Army, to lead the nation in this national counterattack. Those of us privileged to work in this area have an opportunity not only to assist the individual soldiers in the Army and to help the Army as an institution, which is important enough, but also to make a significant contribution to American society as a part of this national effort.

OPERATIONS RESEARCH FOR RISK ANALYSIS EVALUATION

Jerome H. N. Selman
US Army Munitions Command

Dr. Victor Selman
Computer Sciences Corp.

INTRODUCTION

Much attention has been given recently in the Department of Defense to the nature of risk in evaluating major programs, since it has become increasingly clear that the methods and practices used in the acquisition and control of systems and programs have accounted very poorly for risks. Due to the cost growth, time growth, and performance degradation of major weapon systems, the Aerospace Industries Association stated a need for more formal methods of risk assessment. On 31 July 1969, Deputy Secretary of Defense David Packard wrote to the Secretaries of the Army, Navy, and Air Force:

"I would, therefore, like each of you to assure that:

Areas of high risk are identified and fully considered;

Formal risk analysis of each program is made;

Summaries of these are made part of the back-up material for the program."

He later, on 28 May 1970, offered guidance on how risks inherent in new programs can be minimized.

"1. RISK ASSESSMENT. Make a careful assessment of the technical problems involved and a judgment as to how much effort is likely to be necessary in finding a solution that is practical. A careful look at the consequence of failure, even of low risk program elements, is also critical.

2. SYSTEM (& HARDWARE) PROOFING. Perform enough actual (eng.) design and component testing in the conceptual development stage to demonstrate that the technical risks have been eliminated or reduced to a reasonable level. Component or complete system prototyping, or back-up development, are examples of this. Pilot studies, feasibility studies of competing approaches are other examples.

3. TRADE-OFFS (RISK AVOIDANCE). Consider trade-offs not only at the beginning of the program but continually throughout the development stage; program risk and cost are dependent on practical trade-offs between stated operating requirements and engineering design."

Preceding page blank

Risk analysis itself, or risk assessment, requires some clarification as to our current usage. The necessity for making decisions in the face of uncertainty is part of our life style. We must act without knowing precisely the consequences that will result from the action. To deal with large scale system problems, as well as personal problems, we must develop a theoretical structure for decision making that includes uncertainty. The problem of describing uncertainty has perplexed philosophers for centuries: Aristotle, Pascal, Fermat, Adam Smith, Bernoulli, Bayes, LaPlace, Boole, Markov, DeFinetti. Pascal and Fermat laid the mathematical foundations of probability theory 300 years ago.

Conceptually, risk emerges from the fact that some of the information which is pertinent to a decision can, at times, be known in the form of a fuzzy probability distribution. The resulting possibility of deviations from any estimate of the outcome spectrum is the basic phenomenon which gives rise to risk. In the risk analysis methodology Knight's* distinction is made between "risk" and "uncertainty." Technical uncertainties fall into two main categories:

The things you know you don't know when you start the program and for which allowances can be made.

The things you don't know you don't know and for which one is unable to plan--the Unk-Unks!

Within the six phases of weapon systems development---mission concept, system definition, system design, system development, fabrication, assembly and test, operation and support---24 individual steps are identifiable (19 in the development stages). At some point after production and deployment, technical uncertainty approaches zero. The nature and sequence of uncertainties are practically independent of the particular type of system of subsystem under development.

A thorough contract definition reveals many unknowns requiring special attention and plans can be made early to cope with them. For the unknown-unknowns, the "unk unks", the use of paper design and analytical studies alone will not permit high levels of confidence in forecasts of cost, performance and schedule. Unknown factors include actual hardware performance, mutual interference, combinations of environmental conditions, sequence-sensitive effects and the overall impact of combined stress. Planning

*F. H. Knight, Risk, Uncertainty & Profit, Houghton Mifflin Co., Boston, 1921. Risk refers to situations where the outcome is not certain, but where the probabilities of the alternative outcomes are known, or can at least be estimated with some rationale. Uncertainty is where the unknown outcomes cannot be predicted in probabilistic terms, i.e., it refers to contingencies against which one cannot protect himself on insurance principles.

continuity---funding limitations, change of threat, availability of new technology and changes in availability of government furnished equipment---also contribute to the unk-unks. This schema is shown in Fig. 1 with comments on unk-unks.

Fundamental to the risk analysis methodology is the trade-offs in the 3 space of performance-cost-schedule. Cost and schedule growth are integrally related to the performance envelope: As cited by Dr. Baron George of AIA to the Congressional committee in 1969 and shown in Fig. 2.

The factors which influence risk, both in resource limitations and management practices are tabulated in Fig. 3.

In a speech by Major General John R. Guthrie on risk analysis to the Project Managers' Conference, February 1970, he commented that the "most rudimentary sort of good risk analysis might have enabled us to avoid most of the pitfalls we have encountered. By rudimentary I mean---did we identify those items which were new and identify the impact on overall systems performance if that particular component or subsystem were to experience difficulty? I think the M60A1E2 tank is a classic example." MG Guthrie detailed four steps of risk analysis as follows on Fig. 4. MG Guthrie concluded that "The making of a technical risk assessment is extremely difficult, and based upon the AIA study, we have not really been successful in coming up with a good criteria for use at each major decision point."

Several approaches to decision making have been investigated with specific objective of incorporating humans into systems and of making systems more meaningful to people. Analyses have been performed at different levels from the Politico-economic studies in DoD to studies related to weapon system development, to Quantitative subsystem-oriented OR research studies often of an academic approach, to socio-technological studies with or without behavioral science inputs. One such approach is known as decision analysis---a combination of philosophy, methodology, practice and application useful in the formal introduction of logic and preferences to decisions, shown in Fig. 5. The assertions inherent in this analysis (per R. A. Howard) are as follows in Fig. 6.

The decision process itself has come under scrutiny. The most difficult part of the process being the requirement that the decision maker specify his own preferences. A categorization of the decision process is attempted illustrating what objective and subjective probability means relative to observations.

In the objective probability sense, probability is a physical characteristic of an object---weight, volume, loudness. A coin therefore "has" a probability of falling heads on any toss, and to

measure this probability would only require a large number of tosses. The other approach, the subjective approach, considers probability as a measure of the state of knowledge about phenomena rather than about the phenomena themselves. One assigns a probability of heads on the next ross of a coin based upon all the knowledge available about the coin. This distinction in approach though seemingly trivial, is the key to the power of risk analysis. This subjective view of probability is not new. It was clearly proposed by Bayes and LaPlace 200 years ago.

DEFINITION OF RISK ANALYSIS

Risk analysis is the name tag given to a broad spectrum of operations research/systems analysis techniques (both qualitative and quantitative) for analyzing, quantifying, understanding and possibly reducing the uncertainty inherent in the realization of time-cost and/or performance goals of large scale systems. Some descriptions of Decision Analysis are given in Appendix I, with an early chronology of provocative quotes dealing with Decision/Risk Analysis, Appendix II.

WORKING DEFINITION OF RISK ANALYSIS

A working definition of risk analysis might be "a systems analysis or operations research approach to risk" which implies that the goal is to identify the risk areas, reduce or eliminate the risks, and improve the chances of successful accomplishment of the mission. In the Department of Defense the term means the identification of the uncertainties involved with the Time (schedule)/Cost/Technical Performance (quality) measures of the project. Decision Risk Analysis is the method whereby the uncertainty measures of three-dimensional space are traded off to find an optimal, or satisfactory, alternative. The output of decision analysis is a quantitative assessment of which alternative(s) should be selected. A number of techniques of operations research, systems analysis, and management science can be used in forming a decision analysis model of a complex system. The Delphi procedures (for group consensus of experts) and the Standard Gamble or Lottery technique are often used for encoding judgment into (subjective) probability distributions associated with uncertain outcomes for 3-space.

Another useful technique is the Decision Tree methodology which produces useful schematic summaries for structuring and evaluating the alternatives appropriate to a set of circumstances involved with a sequence of decisions. A "tree" is constructed by enumerating and tracking through from start to finish the outcomes of each possible decision that can be made at decision points along the way. The payoff of each route (branch, sequence of decisions, scenario) through the tree is calculated along with the system risk in 3-space.

Probability theory and utility theory are used to tie this all together and determine the expected project payoff and associated risk. Several computer programs have proven useful in carrying out Decision/Risk Analysis in practice.

AMC DECISION RISK ANALYSIS GUIDELINES

Define just what the problem is (this may be a major effort in some instances).

Establish alternatives with their appropriate terminal milestones. (It is important to ferret out all possibilities. An alternative should not be ignored because it does not appear to be a likely future choice).

Lay out all the possible chains of events leading to the terminal milestone for each alternative.

Determine the possible outcomes at the terminal milestone for each alternative in terms such as time, cost and/or performance.

Assess the probability of achieving each of these outcomes. (One should place emphasis on quantifying the uncertainty in those events shown by sensitivity analyses to be driving forces. This effort may be facilitated by developing probabilistic performance models relating component performance to overall performance and utilizing certain computer models which relate total time and cost distributions associated with the terminal milestone to the time and cost distributions of events leading to the terminal milestone).

Conduct trade-off analyses to provide the basis for selecting a preferred alternative.

Determine the sensitivity of this selection to variations in trade-off criteria and sensitive events.

Present the final study to the decision maker in a concise logical fashion emphasing the rationale behind the selection of the preferred alternative. (It is important to highlight the events to which the outcome of each alternative is sensitive.

EXTENSION TO DECISION MAKING

In decision making the alternatives are characterized by multiple attributes (or properties). Weapon systems may be characterized by vulnerability, reliability. cost, yield and other such diversely measured attributes. How is a decisionmaker to choose from among complex alternatives? Clearly, decisionmakers do choose--decisions involving very complex alternatives are made all the time. This is not

to say, though, that these decisions could not be improved. Most decisionmakers in such situations would like a method that would help them process the attribute-value information for each alternative.

Various methods have been proposed to help the decisionmaker with multiple-attribute decisionmaking. These range from techniques which consider all attributes at once to those which consider just single attributes, or proceed sequentially over single attributes, including Dominance, "Satisficing" Maximim Maximax, Lexicography, Additive Weighting, Effectiveness Index, Utility Theory, Trade-offs Non-metric Scaling, and an heuristic combination or sequence of techniques.

Appendix III Risk Analysis Technique gives some useful decision rules in the certainty-risk-uncertainty framework.

APPENDIX I

SOME DEFINITIONS OF DECISION ANALYSIS*

Roberts, Harry V., "The New Business Statistics", The Journal of Business, VOL 33, Jan. 1960:

"...decision theory is a theory of rational behavior in the face of uncertainty;..."

Howard, Ronald A., "The Foundations of Decision Analysis", IEEE Transactions on Systems Science and Cybernetics, VOL SSC-4, No. 3, Sept. 1968:

"...Decision analysis is a term that describes a combination of philosophy, methodology, practice and application useful in the formal introduction of logic and preferences to the decisions of the world. ..."

North, D. Warner, "A Tutorial Introduction to Decision Theory", IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 3, Sept. 1968:

"...Decision theory is a way of formalizing common sense..."

Wilson, Robert B., "Decision Analysis in a Corporation", IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 3, Sept. 1968:

"...Decision analysis is a methodology for analyzing complex decision problems..."

---

*From "A Categorized Bibliography on Decision and Risk Analysis" by . Kahn, Information Sciences Laboratory, General Electric, Schenectady, New York (also applicable to Appendix II).

APPENDIX II

SOME PROVOCATIVE QUOTES DEALING WITH DECISION AND RISK ANALYSIS

(Arranged by Date)

Business Week, "Thew New Business Statistics", March 24, 1962:

"Breakthrough...Bayesian decision theory...opens up just about any business problem to mathematical analysis."

"...The predictions of xperienced sales executives were combined into a probability curve of potential sales. When these opinions were fed into the computer along with cost data, the computer told Du Pont what size plant to build. Du Pont won't say whether it is following the Bayesian conclusion in this case. But is is certainly encouraging its statisticians to persevere in their work."

"...Says Harry V. Roberts, professor of statistics in the University of Chicago's Business School, for the first time we have a bridge between the businessman and the statistician."

"...Schlaifer thinks that the first new frontier that the Bayesian technique will crack will be marketing research."

"...The University of Chicago's Roberts predicts that '20 years from now Bayesian analysis will be the standard way to solve many business problems.' The way things are going, it may come much faster than that."

Green, Paul E., "Bayesian Decision Theory in Pricing Strategy", Journal of Marketing, 27, January 1963:

"However, in contrast to the large number of theoretical contributions being made to decision theory in general and Bayesian statistics in particular, reported applications of thes procedures to real-world problem situations have been rather meager..."

Hertz, David B., "Risk Analysis in Capital Investment", Harvard Business Review, 42, January/February 1964:

"The enthusiasm with which managements exposed to this approach have received it suggests that it may have wide application..."

Heany, Donald F., "Is TIMS talking to Itself?", Management Science, Vol. 12, No. 4, December 1965:

"a gap currently exists between 'managers' on the one hand and many 'scientists' doing research in and on business under the banner of Management Science..."

Howard, Ronald A., "Bayesian Decision Models for System Engineering", IEEE Transactions on Systems Science and Cybernetics, VOL SSC-1, No. 1, November 1965:

"...There is virtually no limit to the potential application of this (decision) theory at every level from power system operation to voltage divider design. Wherever the system engineer encounters decisions in the face of uncertainty, he can now enjoy the aid of a conceptually satisfying and practically powerful inference theory."

Howard, Ronald A., "Information Value Theory", IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-2, No. 1, August 1966:

"If information value theory and associated decision theoretic structures do not in the future occupy a large part of the education of engineers, then the engineering profession will find that its traditional role of managing scientific and economic resources for the benefit of man has been forfeited to another profession."

Howard, Ronald A., "Value of Information Lotteries", IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-3, No. 1, June 1967:

"...we have shown that even an elementary problem of this type may be far from trivial in the familiarity with probabilistic operations required to derive the results one would like to examine. In view of this observation, it is not surprising that so few managers and engineers make use of formal decision models in decisionmaking."

"Yet it is inevitable that in the future both technical and managerial decisionmakers will employ formal logical methods in decision-making. The transition probably will be painful."

Howard, Ronald A., "The Foundations of Decision Analysis", IEEE Transactions on Decision Analysis", IEEE Transactions on Systems Science and Cyber netics, Vol. SSC-4, No. 3, Sept. 1968:

"...There was a time less than a decade ago when suggesting that decision theory had practical application evoked only doubtful comment from decision makers. The past five years have shown not only that decision theory has important practical application, but also that it can form the basis for a new professional discipline, the discipline of decision analysis..."

"...The last few years have seen decision analysis grow from a theorists's toy to an important ally of the decision maker. Significant applications have ranged from the desirability of kidney transplants through electric power system planning to the development of policies for space exploration. No one can say when the limits of this revolution will be reached. Whether the limits even exist depends more on man's psychology than on this intellect."

Spetzler, Carl S., "The Development of a Corporate Risk Policy for Capital Investment Decisions", <u>IEEE Transactions on Systems Science and Cybernetics,</u> VOL. SSC-4, No. 3, Sept. 1968:

"It is still too early to report on the company's experience of using a risk policy. However, the company's willingness to commit the required resources first to carry out this project and now to continue with application indicated a belief in the approach."

"Some of the major effects of the study on the company have been educational..."

DECISION RULES USING CERTAINTY-RISK-UNCERTAINTY CLASSIFICATION

DECISION-MAKING UNDER CERTAINTY: Given a set of possible alternatives A, choose one (or all) of those alternatives which maximize (or minimize) some given index, such as maximum value or maximum utility. The bulk of formal theory in economics, psychology, and management science falls here. Linear programming solutions are often used due to their nexus with the Theory of Games.

DECISION-MAKING UNDER RISK: Each action may result in more than one outcome depending upon the state of nature, where each state of nature has a known, or presumed known probability.

1. Maximum Expected Value Criterion. Select that alternative, a_i, with the maximum gain (or minimum loss):

$$\max_i E(a_i) = \max_i \sum_{j=1}^{n} v_{ij}(p_j); \text{ where } j = 1, 2, \ldots n$$

2. Maximum Subjective Expected Utility Criterion. Same as above, but with utilities instead of values:

$$\max_i E(a_i) = \max_i \sum_{j=1}^{n} u_{ij}(p_j)$$

3. Most Probable Future Principle Criterion. Use that state S_j for which p_j is a maximum; select that alternative a_i which has the maximum v_{ij} or u_{ij}. In practice many decisions are made this way.

4. Expectation-Variance Principle. Considers both the expected value or utility) and the variance of the alternatives. If two alternatives have the same expected utility, select the alternative with the smaller variance. If two alternatives have the same variance, pick the a_i with the larger expected value(or utility).

5. Simon-March "Satisficing" Hypothesis. "Most human decision making, whether organizational or individual, is concerned with the discovery and selection of satisfactory alternatives; only in exceptional cases is it concerned with the discovery and selection of optimal alternatives."

6. Bayes Decision Criterion. This rule is similar to LaPlace in using p_j's; however, the probabilities associated with the possible future states of nature are subjectively estimated by the decision maker, and not necessarily equiprobable. Hopefully, the subjective estimates will consider hard data, intuition, judgment and brinkmanship. The Expected Value or Utility criterion is used to select the alternative, the $\max_i E(a_i)$.

DECISION-MAKING UNDER COMPLETE UNCERTAINTY: The following rules do not depend upon the assignment of probabilities to the states of nature; therefore, they are non-stochastic decision rules. Each action may result in more than one outcome, depending upon the state of nature; but each state of nature has an unknown probability.

1. The LaPlace Criterion of Insufficient Reason. If the probabilities of the different possible states of nature are unknown, or meaningless, this complete "ignorance" is expressed in mathematical terms by assuming that all the probabilities are equal. This follows from the formulation of J. Bernoulli who stated that if there is no evidence that one event (state of nature) is more likely to occur than another, the events should be judged equally likely. LaPlace said that the arithmetic average of the utilities should be taken. The LaPlace Decision Rule is to choose that action alternative a_i which gives

$$\boxed{\operatorname*{Max}_{i} \frac{1}{n} \sum_{j=1}^{n} u_{ij}}$$

2. The Wald "Sure-Thing" Principle of Minimax (Maximin). This game-theoretic rule from the two-person-zero-sum game ,(2POS), has been called pessimistic, conservative and rational. It protects against a high loss; the best-of-the-worst, or worst-of-the-best kind of payoffs. The rule states that we should find the smallest value for each alternative, and then choose that alternative with the maximum of the minimum values, the maximin (or for loss matrices, the minimax). This may be expressed mathematically as

$$\boxed{\operatorname*{Max}_{i} \operatorname*{Min}_{j} u_{ij}}$$

3. The Maximax Criterion. Find the largest value for each alternative a_i, and then choose that alternative with the maximum of the maximum values. the maximax, or $\max\limits_{i} \max\limits_{j} u_{ij}$.

4. The Hurwicz Alpha-Criterion of Optimism/Pessimism. L. Hurwicz combined the minimax and the maximax criteria in a convex linear combination, permitting all levels of optimism with his coefficient of optimism, alpha, which can vary between zero and unity. When alpha is zero, Hurwicz reduces to the Minimax Rule; when alpha is unity, the Maximax Rule. Choose that alternative which gives

$$\boxed{\operatorname*{Max}_{i} \left(\alpha \operatorname*{Max}_{j} u_{ij} + (1-\alpha) \operatorname*{Min}_{j} u_{ij}\right)}$$

5. The Savage Minimax Regret Criterion. L.J.Savage suggested this criterion obtained by applying the minimax criterion not to the original payoff matrix, but to a new matrix, known as the regret matrix. The entries in the regret matrix are obtained from those in the payoff matrix by subtracting from each of the entries the maximum entry in the same column. Define the regret (negative) matrix, r_{ij}, equa · $(u_{ij} - \max u_{kj})$, where r_{ij} measures the regret between the payoff which is actually obtained and the payoff which would have been obtained if the true state of nature, p_j, had been known. Applying the Minimax Rule choose that alternative a_i for which the min r_{ij} is maximized, $\max_j \min_i r_{ij}$.

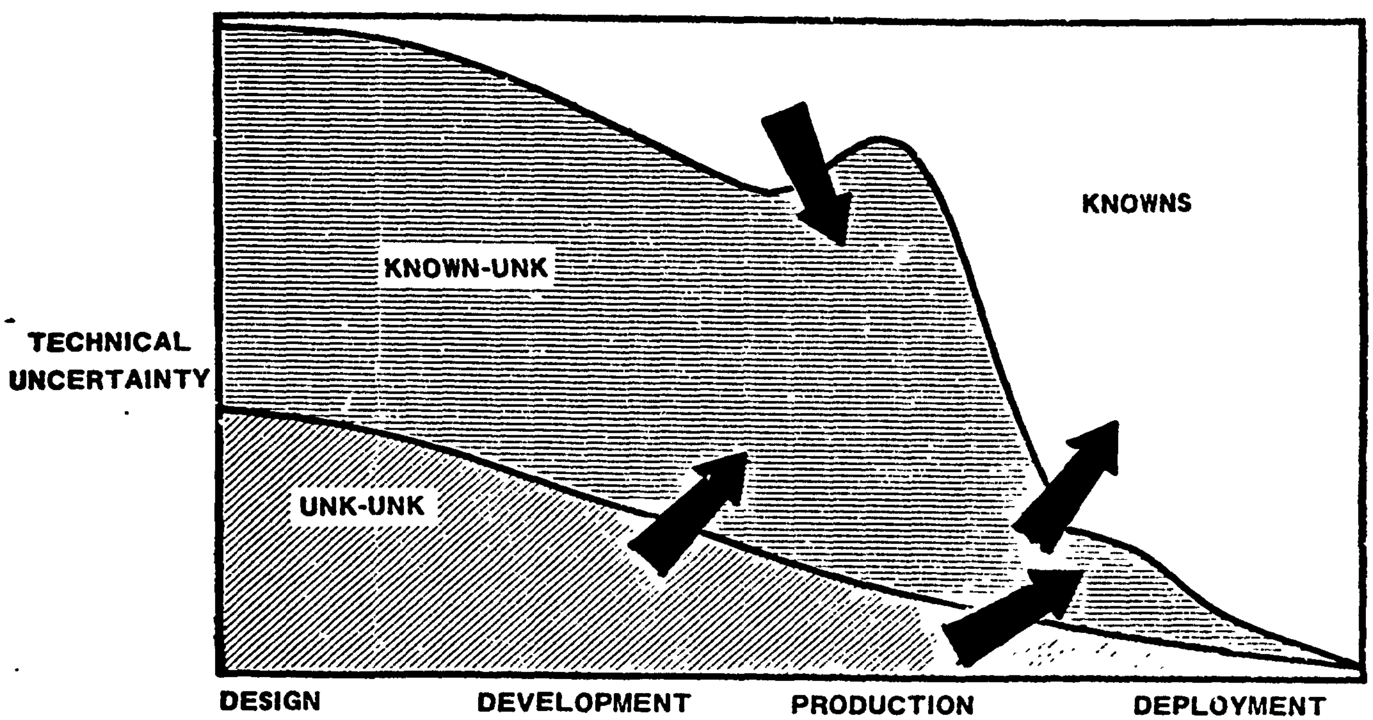
The above non-stochastic decision rules which are based upon "ignorance" of the states of nature to optimism, to regret all influence the final decision. In the literature is the following decision matrix where the preferred action in all cases depends upon the decision rule used!

$$\begin{array}{c} A1 \\ A2 \\ A3 \\ A4 \end{array} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 3 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 2 & 2 & 0 & 1 \end{bmatrix} \begin{array}{l} \text{minimax loss} \\ \text{minimax regret} \\ \text{Hurwicz for } \alpha > 1/4 \\ \text{insufficient reason} \end{array}$$

RISK ANALYSIS TECHNIQUES

DECISION PROBLEM UNDER CERTAINTY (DPUC)	DETERMINISTIC OR TOOLS (E.G., LINEAR PROGRAMMING)
DECISION PROBLEM UNDER RISK (DPUR)	PROBABILITY THEORY, UTILITY THEORY
DECISION PROBLEM UNDER UNCERTAINTY (DPUU)	GAME THEORY (CONSTANT SUM, NON-CONSTANT SUM AND METAGAMES), INFORMATION THEORY
DECISION PROBLEM UNDER RISK AND UNCERTAINTY (DPUR/U) - CONFLICT	DECISION THEORY, UTILITY THEORY

UNKNOWNS TO KNOWNS TRANSITION



Mr. Holifield. You would not mind my having a little pun here that that unk unk becomes a flunk flunk?

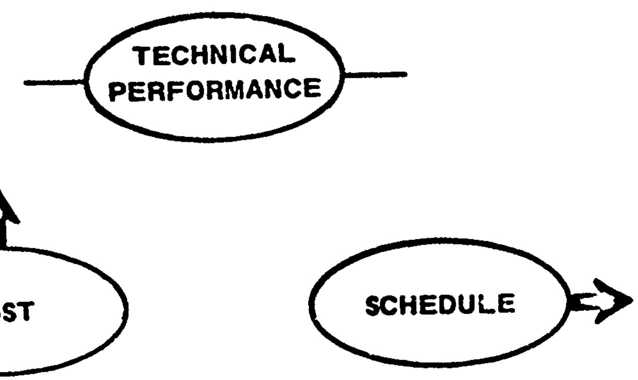
Dr. George. Not at all. Very good. May I quote you?

Mr. Holifield. Yes

Hearings before a subcommittee of the committee on Government Operations House of Representatives, June 19, 24, 25 and 27, 1969, Subject: Government Procurement and Contracting (Part 9).

FIGURE 2

PERFORMANCE-COST-SCHEDULE TRADE-OFF



TECHNICAL RISK ASSESSMENT IS CRUCIAL

DR. GEORGE. THIS IS A RATHER SIMPLE CHART, BUT A VERY IMPORTANT ONE. IT IS REALLY THE HEART OF THE WHOLE PROBLEM!

AT EACH OF THOSE DECISION POINTS I SHOWED IN THE PREVIOUS CHART, YOU HAVE TO MAKE A PERFORMANCE, COST AND SCHEDULE TRADEOFF, REPRESENTED BY THESE THREE COLORED AREAS. IF I ELECT TO HOLD THE TECHNICAL PERFORMANCE FIXED IN THE FACE OF TECHNICAL UNCERTAINTIES, UNKNOWN UNKNOWNS, AS THE CHAIRMAN HAS REFERRED TO THEM, THEN IT IS INEVITABLE THAT THE COST WILL HAVE TO GO UP AND PROBABLY THE SCHEDULE WILL GO TO THE RIGHT. THIS MEANS THAT THE TECHNICAL RISK ASSESSMENT IS A REAL CRUCIAL MATTER IF YOU ARE TO MAKE THIS TRADEOFF IN THE BEST PROFESSIONAL MANNER.

FIGURE 3

FACTORS WHICH VARY THE RISK

- RESOURCE LIMITATIONS:
  - LACK OF DEFINED REQUIREMENTS
  - INSUFFICIENT QUALIFIED PERSONNEL
  - LACK OF KNOWLEDGE
  - LACK OF PROVEN COMPONENTS
  - LACK OF DESIGN MARGINS
  - LACK OF TIME
  - INSUFFICIENT FUNDS
- MANAGEMENT PRACTICES:

  EXTENT TO WHICH PROVEN PRACTICES ARE USED IN THE AREAS OF

  - ENGINEERING
  - MANUFACTURING
  - QUALITY CONTROL
  - PROGRAM

FIGURE 4

FOUR STEPS OF RISK ANALYSIS:

- PROBLEM FORMULATION - CLEARLY STATING THE OBJECTIVES; DEFINING THE ISSUES, IN PARTICULAR THOSE THAT ARE INHERENTLY SUBJECT TO UNCERTAINTY

- SEARCH FOR RELEVANT DATA - HOW MUCH IS KNOWN TO BE FACT AND WHAT IS VAGUE OR SUSPECT. WHAT HISTORY IS AVAILABLE ON SOMEWHAT SIMILAR SYSTEMS IN THE PAST? WHAT PITFALLS, WITH RESPECT TO PROJECT COSTS, TIME AND PERFORMANCE, WERE ENCOUNTERED?

- MODEL BUILDING - THIS PROCESS ATTEMPTS TO DESCRIBE THE PHYSICAL SITUATION UNDER STUDY IN MATHEMATICAL AND PROBABILISTIC TERMS WITH RESPECT TO THE USUAL OBJECTIVES OF (A) TIME TO COMPLETION OF PROJECT; (B) PROJECT COSTS; AND (C) GOALS ON PERFORMANCE. THIS MODEL WILL BE USED TO AID THE DECISION MAKER IN DECIDING THE RELATIVE CAPABILITY OF MANY COMPETING ALTERNATIVES TO MEET THE STATED OBJECTIVES

- CRITERIA TO DISCRIMINATE BETWEEN VARIOUS ALTERNATIVES - WHAT DIFFERENCE IN PERFORMANCE BETWEEN COMPETING ALTERNATIVES IS CONSIDERED SUFFICIENTLY LARGE TO BE CONSIDERED SIGNIFICANT AND STILL HAVE UTILITY OR MILITARY WORTH TO THE USER

FIGURE 5

THE DECISION ANALYSIS CYCLE

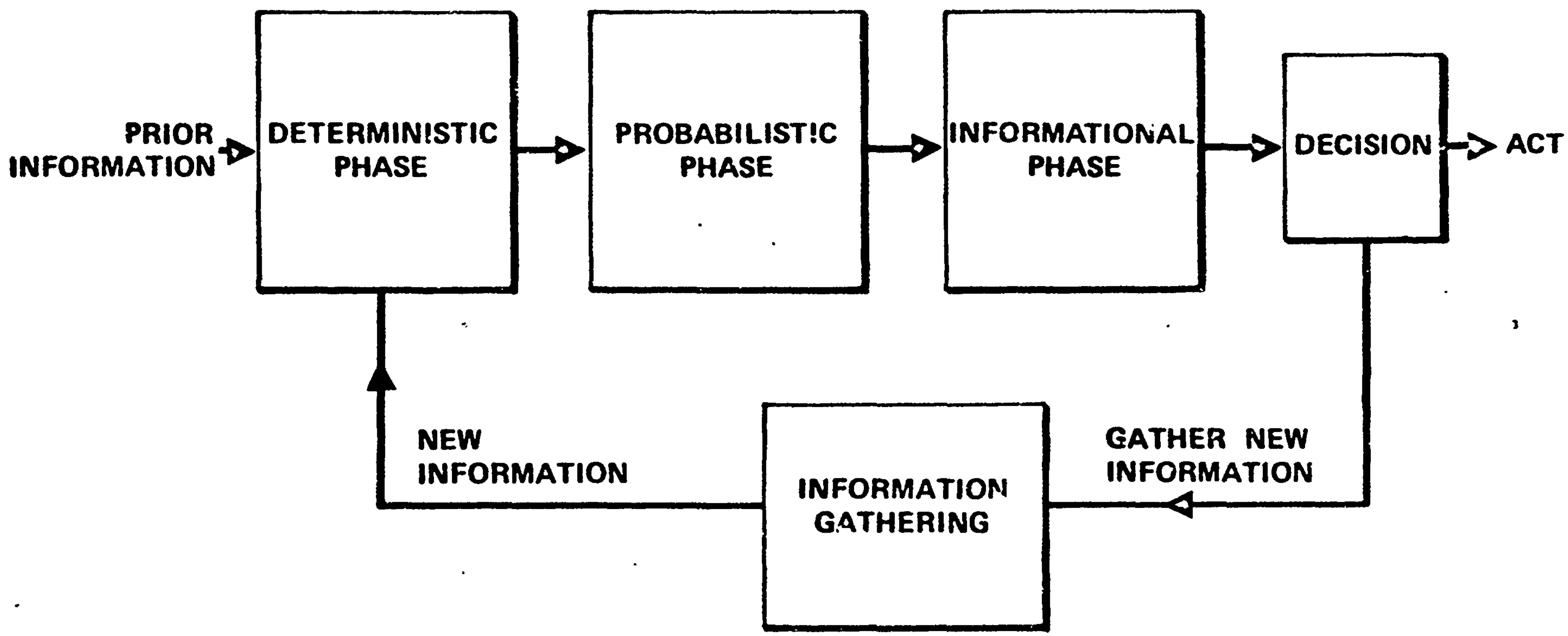


FIGURE 6

RISK ANALYSIS

ASSERTIONS*

- THE PROCESS OF DECISION-MAKING IS AT THE HEART OF MOST TECHNICAL, BUSINESS AND GOVERNMENTAL (RISK ASSESSMENT) PROBLEMS
- DECISION-MAKING REQUIRES THE STUDY OF UNCERTAINTY
- UNCERTAINTY CAN ONLY BE STUDIED FORMALLY THROUGH PROBABILITY THEORY
- PROBABILITY IS A STATE OF MIND, NOT OF THINGS
- ALL PRIOR EXPERIENCE MUST BE USED IN ASSESSING PROBABILITIES
- DECISION-MAKING REQUIRES THE ASSESSMENT OF VALUES AS WELL AS PROBABILITIES
- DECISIONS CAN ONLY BE MADE WHEN A CRITERION IS ESTABLISHED FOR CHOOSING AMONG ALTERNATIVES
- THE IMPLICATIONS OF THE PRESENT DECISION FOR THE FUTURE MUST BE CONSIDERED
- WE MUST DISTINGUISH BETWEEN A GOOD DECISION AND A GOOD OUTCOME

*R. A. HOWARD, THE SCIENCE OF DECISION-MAKING. REPRINTED BY THE JOINT ENGINEERING-ECONOMIC SYSTEMS PROGRAM, STANFORD RESEARCH INSTITUTE

CLIENT CENTERED RISK ANALYSIS

By: Mr. Lawrence L. Rosendorf
Picatinny Arsenal

In his recent best seller "The Greening of America", Charles Reich has advocated that Americans adopt a new consciousness to solve the problems of society. It will be the thesis of this paper that the risk analyst must also adapt a new ·onsciousness to solve his problems.

To lay the groundwork for this thesis, we will have to explore:

a. The role of the risk analyst.

b. His location in the organizational structure.

c. His conception of human nature.

The risk analyst is most often identified as a person with an Operations Research/System Analysis background who can be called into a project and provide an objective assessment of project alternatives, compatibility and uncertainties of schedules, funding and performance. His key "raison-d'etre" is therefore to gather data, question the data and coalate it (thru the use of models) into a means of communicating with management such that he can aid in optimizing the decision process. Very often, the risk analyst is called upon on short notice, (e.g., before a key briefing) to provide these assessments. He must then tend to the difficult task of acquiring, screening and iterating all of the data before putting it into his model for evaluation.

I believe that all Operations Research/Systems Analyst/R.A. are sensitive to the importance of good input data into their models and the subordinate importance of the sophistication of their models.

As a reminder of this I would like to reiterate two of the pitfalls of systems analysis that had been originally stated by Herman Kahn. These are modelism and overambition. Modelism is roughly defined as the event that occurs when the systems analyst becomes more involved in his model and model building than in solving the real world problem, and overambition is the tendency of systems analysts to try to cover too much of one problem area in any reasonable fashion. As an interesting sidelight to this, Phillip Greene, in his book "Deadly Logic", severly chides Herman Kahn for exhibiting these two same pitfalls in Kahn's work, "On Thermonuclear War."

Now let us examine for a moment, the location of the risk analyst in the organizational structure.

Preceding page blank

The risk analyst is in a large sense today, in the same situation that the Quality Assurance analyst was several decades ago. He is often feared, avoided, circumvented and detoured. He is generally looked upon by project personnel as a nuisance who will take their valuable time, uncover no significant facts and in general, merely do what they have been doing anyway. He is usually provided with necessary data only when project personnel are ordered to "cooperate" from some higher level of organizational authority. Even then project personnel enter into the risk analysis study with a defensive, "why me?" attitude.

For reasons of objectivity and independence, the risk analyst is set aside organizationally from project personnel. This is to provide him with the necessary freedom to be fully objective in his assessments of risk in much the same way that quality assurance personnel are separate or the comptroller is separate. These are the people that are to provide the necessary checks and balances to the organization and moderate the optimism, extravagance or whatever, of project personnel. The risk analyst is often thought of in elitest terms in the organization. He is expected to be an alter ego of the top management, striving always for truth and objectivity with respect to the organizational mission.

It is my thesis, that placing him in such an organizational structure, where he is independent of the lines of authority of project personnel is mandatory and yet also untenable, that is unless the risk analyst adopts a new consciousness. Therein lies the paradox. The case for the separation of the risk analyst from project personnel can be stated very strongly, and will not be elaborated on in this paper. Why the situation is untenable can be realized by examining some of the data gathering techniques at the disposal of the risk analyst. The list can be shown to include the 4 basic methods of data acquisition and 11 other possible combinations of these 4 taken 2, 3, or 4 at a time. (Table 1). This list is not necessarily all inclusive but represents the major sources of data acquisition.

TABLE 1

DATA ACQUISITION TECHNIQUES

a_1	:	Question the engineers most closely related to the project.	
a_2	:	Assemblc a group of non-project related experts in the areas of technology being considered by the project.	
a_3	:	Utilize a data base or data bank (i.e., report, files, books, computer stored info. etc.).	
a_4	:	Employ DELPHI techniques, utilizing both project and non-project related personnel, in an iterative, questionnaire-feed back process.	
a_5	:	Union of a_1 and a_2	
a_6	:	Union of a_1 and a_3	
a_7	:	Union of a_1 and a_4	$\binom{4}{2}$
a_8	:	Union of a_2 and a_3	
a_9	:	Union of a_2 and a_4	
a_{10}	:	Union of a_3 and a_4	
a_{11}	:	Union of a_1, a_2, a_3	
a_{12}	:	Union of a_1, a_2, a_4	$\binom{4}{3}$
a_{13}	:	Union of a_2, a_3, a_4	
a_{14}	:	Union of a_1, a_3, a_4	
a_{15}	:	Union of a_1, a_3, a_4	$\binom{4}{4}$

The risk analyst is generally called into programs on an ad-hoc basis. He must therefore educate himself quickly, acquire the data, evaluate the data with his models and present some recommendations based on these. If he turns to project personnel for inputs, he must be prepared to question the validity of these inputs, and to moderate the possible optimism of these inputs based on his own background or experiences in similar programs, schedules, funds or state of technology. This is asking a great deal from a risk analyst as he is often not at all familiar with the technology or program being analyzed. Most people tend to resolve this by saying:

a. The risk analyst must be a man of great individual quality, education, background, etc.

b. The risk analysis organization should have a permanent staff of experts who can be assembled on an as required basis for a given risk analysis. or

c. The risk analyst should assemble an ad-hoc team of experts who are non project related, and may not reside within his organization, to aid him in an objective evaluation of the uncertainties in the project.

I submit the there are very few people in a large diversified organization who can satisfy the job description of the risk analyst (in a. above) as having the breadth and depth of knowledge required to analyze a large portion of the commodities within any large diversified organization's mission responsibility and then evaluate the consequences of the analyses thru the use of O.R. type models. William Souder, in a paper titled, "The Validity of Subjective Probability of Success Forecasts by R&D Project Managers" in IEEE Trans. Engineering Management, Vol. EM-16, February 1969, has attempted to measure the predictive validity and consistency of forecasts to be obtained from R&D project managers, in an attempt to aid management in the early identification of eventually failing projects. In discussing some possible reasons for the non-correlation of one particular projects' optimistic probability of success with that projects eventual failure Souder's states:

"Project 10 gave every indication....of being a future success. Yet the project ended in failure 24 months after it started. Latent technical problems that became apparent during the second year caused management to abandon the effort. A postmortem on the project showed that these technical problems were, in fact, recognized by the project manager well in advance of the time that they were communicated to upper management! Moreover, the postmortem showed that some of the chemists reporting to the project manager were concerned about the severity of these problems during the period covered by the experiment. They were, in fact, discussing these problems among themselves, but not reporting them outside the group. Only optimistic assessments of future potentials were given by the respondent during the experiment.

In holding back information that turned out to be vital to the project, the project manager was not necessarily purposely attempting to conceal the true facts. A complex set of environmental factors, organizational factors, and respondent personality factors appeared to be at least partially responsible for this obscuring of salient problems. For example, this particular product area had been publicly cited by top management as a possible spectacular one for the near future. Therefore, there was a reluctance to report any bad news."

I would also submit that it would be organizationally very inefficient to have a staff of full time independent, non project related technical experts in the Risk Analysis Office. Most of these experts could be better utilized in the project/system engineering directorates, spending their time in the resolution of technical problems as members of the project team.

The third alternative therefore seems to make the most sense. A team of experts can be assembled from either within the organization or from outside consultants on an ad hoc basis, to provide assessment of the risks, based on the approach described by the project personnel. The risk analyst can serve as a catalyst in stimulating and encouraging dialogue between these two groups. Once again, the project people will probably be defensive and feel threatened by the criticism if this is not handled properly.

It is my opinion, that the central problem faced by the risk analyst lies in his obtaining valid input for his model. Very often, when faced with the prospects of confronting a "hostile" project person, the risk analyst is forced to recede back to his comfortable world of modelism in an attempt to mathematize those aspects of the problem for which he lacks data or real information. His data gathering techniques may be confined to use of a data bank or perhaps a Delphic approach.

The Delphic Approach, as many of you are aware, was designed to gather information and possibly arrive at a consensus while eliminating many of the undesirable effects of group interaction.

Soloman Asch conducted a most provocative experiment which illustrates the effects of "Opinions and Social Pressures" in his paper of that same name in Scientific American, November 1955.

DELPHI as I am sure you are aware, has the disadvantages of; substantial time delays for repeated iterations plus a heavy reliance on communications based solely on narration and numbers. I tend to feel that the use of DELPHIC approaches in accumulating data is analogous to designing a car with jacks under every wheel. In both cases we are circumventing the main problems, interpersonal-communications in the former and improved tire or road design in the latter, and in both instances we will pay disproportionately for the services rendered.

The basic problem, in terms of organizational interests, is one of possible conflict resolution between the risk analyst and project personnel. Two techniques I would propose for the reduction or resolution of this conflict are both well known in the fields of psychology and social psychology. There are (a) Appeal to a superordinate goal and (b) Client centered therapy.

Sherif, in his paper "Experiments in Group Conflict" in Scientific American, November 1956, discusses the idea of appealing to a superordinate goal.

At an isolated summer camp, boys 11 and 12 years of age were divided into two groups and housed in different cabins. None of the boys were aware they were a part of an experiment on group relations.

Tournament of games, such as baseball, touch football, and tug-of-war were arranged to produce friction between the two groups. Good sportsmanship soon evaporated, as each group refused to have anything more to do with individuals in the opposing group. Solidarity increased within each group.

Sherif hypothesized that groups having conflicting aims (i.e., when one can achieve its ends only at the expense of the other) will become hostile to each other even though the groups are composed of normal well-adjusted individuals.

Can harmony be derived between two groups in conflict? The theory was then tested to reduce friction between them by means of pleasant social contacts between the two groups. Social events, however, only served as opportunities for the rival groups to berate and attack each other.

Sherif then argued, "Just as competition generates friction, working in a common endeavor should promote harmony. The most decisive factor is the existence of "superordinate" goals which have a compelling appeal for both but which neither could achieve without the other." A series of situations were created to test this hypothesis. Gradually the cooperative acts reduced friction and conflict and the two groups became more friendly, whereby in the end, they were actively seeking opportunities to mingle.

"What our limited experiments have shown is that the possibilities for achieving harmony are greatly enhanced when groups are brought together to work toward common ends."

Can the Risk Analyst utilize some of the conclusions from the former experiment to dispel the potential conflict between himself and the project personnel? I believe he can. All too often the Risk Analyst enters upon the scene, with a preconceived model or format for accomplishing the analysis. The interactions, feedback and clarification with project

personnel are minimized. Insufficient time is probably spent in clearly delineating the objectives of the Risk Analysis to the project people, or the purported advantages to them of having one accomplished on their project. I believe that by carefully exploring the possible advantages with the project personnel, the Risk Analyst can develop the commonality of "a superordinate goal" thus reducing conflict and defensiveness on the part of the project personnel.

The manner with which the Risk Analyst approaches the study is therefore critical to its eventual success. Let us now consider the attitude of the Risk Analyst with respect to "Human Nature."

I hope you will pardon me for exploring this from a Psychology 101 framework, but I would like to develop a common framework before proceeding with my discussion. In his book "The Organization Man", William H. Whyte, Jr. states:

> "Ever since Newton, scores of natural scientists have stepped out of their area of competence to suggest the possibilities of a science of man, and Erasmus's "Praise of Folly" suggests that even before this some savants had much the same idea. It was an understandable dream for a natural scientist to have. Even Descartes himself was seized with the idea that the discipline of mathematics could be extended to the affairs of man. Eventually he thought, a "Universal Mathematical Science" would solve the problem of society--if only there were sufficient funds and time for the job.
>
> Later others tried the geometric tack: Thomas Hobbes worked out a complete set of algebraic equations to explain ethics. As Laurence Sterne remarked, his equations "plussed or minussed you to heaven or hell...so that none but the expert mathematician would ever be able to settle his accounts with Saint Peter.""

While it has never been really possible to successfully develop a science of man in the sense of a universal law of relativity, applied to human nature, for gross purposes of classification I would like to offer two of the most widely held theories. The first is that of Sigmund Freud and the second is attributed to Carl Rogers.

At the outset I must qualify my presentation by stating that I am not a psychologist by training, but I feel I am a psychologist by interest.

Freud theorized that a person's behavior was driven by forces of subconscious motivation, and that the goal of behavior is to reduce fear. He felt that all behavior is a compromise, and a result of conflict between a person and his social environment. Further he reasoned, since a person is in conflict, all behavior is defensive.

To attempt to summarize Freud's position with respect to a "Science of Man", I believe we could say Freud viewed man in a negative fashion, being driven by instinctual needs and compromising his gratification of these needs due to the presence of the SUPEREGO. All behavior was an attempt by the EGO to allow instinctual energy to manifest itself. Defense mechanisms developed to satisfy the EGO and must be present. In a crude way we can think of Freud's theory as an allocation of resources problem (where the resource is instinctual energy) with an objective function of maximizing instinctual energy or needs, without exceeding the constraints that society imposes on the SUPEREGO.

On the opposite side of the coin is Carl Rogers. Rogers, like Freud, also started with a concept of the organism, but Rogers feels there is a force within people, which he calls, "An Actualizing Tendency." This "Actualizing Tendency" is not an attempt to rid yourself of instinctual energy, as was Freud's argument, but is a biological tendency to fulfill your genetic blueprint. Given that a person is born with this actualizing tendency, the next stage in his development, according to Rogers, takes place when he contacts the environment. Now, the self concepts develops. These are conscious thoughts and feelings you have about yourself that are comprised of preferences and values. These are not innate, but are learned or experienced, and the self concept is largely shaped by whether the organism's behavior is accepted or rejected by others. The final stage in the development is the principle of the self-actualizing tendency. This is the tendency to keep behavior consistent with the self concept. A person has a very high regard for his self concept, and anything that conflicts with his self concept leads to anxiety. Rogers, as opposed to Freud, says that conflict is not inevitable. If the person were accepted without conditions, then conflict would not arise. Freud says that defenses are good and necessary, as they aid in dissipating the instinctual energy. Rogers feels defenses are bad and unnecessary, as they are preventing and restricting you from being your real self. For Freud, defenses reduce conflict; for Rogers, defenses lead to aggression, tension, and hate.

Both Freud and Rogers developed their respective theories through associations with large number of patients in therapy. Freud used an evaluative approach which was directive in nature; that is, he explained to the patient the suspected causes for his problems. Rogers disagreed. He claims that an urge exists within a person to know about himself, to be fully actualized. He developed a technique of non-directive therapy, which he called, "Client Centered Therapy." The basic premise of "Client Centered Therapy" is that if an individual is placed in a situation that is non-threatening, and if we believe that the individual has the capacity to deal constructively with those aspects of the problem when he is aware of them, then the individual will react in a cooperative, non-defensive manner.

The analyst must be careful to protect against: (a) a seeming lack of interest or involvement, which may be interpreted by the client as rejection; and (b) a laissez-faire attitude which might indicate to the client that the analyst does not consider him to be a person of worth.

In Rogers' words:

> "Client-centered therapy is built on two central hypotheses: (1) the individual has within him the capacity, at least latent, to understand the factors in his life that cause him unhappiness and pain, and to reorganize himself in such a way as to overcome those factors; (2) these powers will become effective if the therapist can establish with the client a relationship sufficiently warm, accepting and understanding. .....In this atmosphere of complete psychological security the client can lay himself bare with no danger of being hurt. ....We take this approach because we have found it to be a deeper and more effective method than any interventive procedures we might use to help the individual deal with life."

By my examination of the positions of Freud and Rogers, with respect to human nature, I hope the reader will not infer that I believe the Risk Analyst should practice psychotherapy on his clients. I would propose, however, that the Risk Analyst synthesize some of these approaches in his data gathering and communications activities.

Two studies that dramatically illustrate aspects of client-centered therapy are: (a) "Teacher Expectations for the Disadvantaged," by Robert Rosenthal and Lenore F. Jacobson in Scientific American, April 1968; and (b) "Non-Verbal Communication," by Merhabian.

Robert Rosenthal a social psychologist at Harvard, and Lenore Jacobson a principal of an elementary school in a South San Francisco unified school district, set out to test the validity of the widely held belief that poor children lag in school because they are members of a disadvantaged group.

Teachers in the school were told that a new type of IQ test was to be administered to their students, and that the test is designed predict academic blooming in children. After the tests were administered, the experimenters selected 5 students in each class as academic bloomers by means of a table of random numbers. The names of these students were given to their teachers in a very casual manner at the start of the next semester, as children who could be expected to show unusual intellectual gains in the year ahead. The difference between these children selected from the random number table and the children who were not selected, was solely in the minds of the teachers.

The children were retested four months later and at the end of the school year.

The results indicated strongly that children from whom teachers expected greater intellectual gains showed such gains.

At the end of the academic year the teachers were asked to describe the classroom behavior of their pupils. The children from whom intellectual growth was expected were described as having a better chance of being successful in later life and as being happier, more curious and more interesting than the other children. There was also a tendency for the designated children to be seen as more appealing, better adjusted and more affectionate, and as less in need of social approval.

"How is one to account for the fact that the children who were expected to gain did gain? The first answer that comes to mind is that the teachers must have spent more time with them than with the children of whom nothing was said. This hypothesis seems to be wrong, judging not only from some questions we asked the teachers about the time they spent with their pupils but also from the fact that in a given classroom the more the "spurters" gained in I.Q., the more the other children gained.

Another bit of evidence that the hypothesis is wrong appears in the pattern of the test results. If teachers had talked to the designated children more, which would be the most likely way of investing more time in work with them, one might expect to see the largest gains in verbal intelligence. In actuality the largest gains were in reasoning intelligence.

It would seem that the explanation we are seeking lies in a subtler feature of the interaction of the teacher and her pupils. Her tone of voice, facial expression, touch and posture may be the means by which--probably quite unwittingly--she communicates her expectations to the pupils. Such communication might help the child by changing his conception of himself, his anticipation of his own behavior, his motivation or his cognitive skills. This is an area in which further research is clearly needed."

Merhabian's study may shed some additional light on the communications process. He tried to uncover how much communication takes place at a non-verbal level. In his experiments, experimenters made either positive or negative statements to subjects with happy or unhappy facial expression and tone of voice. What do the subjects actually hear and respond to? He found that in situations of conflict (i.e., facial expression inconsistent with verbal message, or verbal tone inconsistent with verbal message) the subjects reports as to what is being said are determined 55% by tone of voice, 38% of the time by facial expression, and 7% by content.

Even if we question the magnitude of the differences, I'm sure you would grant that non-verbal communication is a very important, often overlooked area.

Now, let's focus our attention back on the Risk Analyst. Can he use any of the techniques discussed above in his contacts with project personnel in the pursuit of data? I believe he can.

To put my argument into the framework of a decision theory problem let us identify as actions each of the previously mentioned 15 data gathering techniques or combinations of techniques; that is:

Action a_1 : Question the engineers closely related to the problem;

Action a_2 : Assemble a group of experts; etc

$\vdots$

a_{15}

These are actions or decision alternatives that the Risk Analyst must decide among so as to obtain data for his study. We will assume, for the moment, that he has free choice to select the optimal data acquisition technique based on his criteria.

As is common in decision analysis, let us identify two states of nature, θ_1 and θ_2. These are the attitudes that the Risk Analyst exhibits toward human nature; θ_1 being a Freudian approach (or directive-evaluative) and θ_2 being a client-centered, or Rogerian approach. Implied in θ_1 is the assumption that the Risk Analyst views people pessimistically and feels more comfortable with models than people. Implied in θ_2 is an attitude of trust and respect for the work of the project people. An attitude of being "client-centered" in your contacts implies; appealing to a superordinate goal and demonstrating this through nuances of both verbal and non-verbal communication.

I realize that θ_1 and θ_2 are not necessarily mutually exclusive, nor are they totally inclusive. It can also be mentioned that my analysis is overambitious and suffers from the defects Kahn has warned against. I concur in the above criticism, but I beg your indulgence a bit further.

Let us now set up a matrix, delineating the actions (i.e., the data gathering techniques), the states of nature, and the expected outcome for each. This is shown below for only five of the 15 possible actions.

	θ_1 : Freudian Approach	θ_2 : Rogerian Approach
a_1 : (Proj. people only supply data)	Alienate Project people. Data supplied unwillingly. Delays, aggravation on both sides. Disbelief of data. Distrust. Neither party fully supports final results.	Project people not threatened by R.A. Data supplied willingly in less time and with full cooperation. Conclusions supported by Project and Risk Personnel.
a_2 : Experts only supply data	Project people left out. Data probably poorer than from Project people. Project people will probably disagree and reject study.	Close to Freudian, except possibly modified slightly by R.A. appealing to superordinate goal with Project people.
a_3 : Data base only	Poor data source. Poor schedule information. No communication with project people or experts.	Same as Freudian
a_4 : Delphic Approach	Slow - limited inputs, high reliance on writing skills for communications. High reliance on adequate questionnaire. May be resented as feeling of stifling of communication. Avoids dominating individual interfering.	Same as Freudian
$a_5 = a_1 \cup a_2$	Project personnel further alienated. Now feel there are two checks on them. Faster data acquisition than in a_1 and probably better data. Still poor cooperation - conflict. Improved "group intelligence."	Project personnel not threatened. Expert advice provided as consulting aid to project. Data supplied is honest and forthright. Full cooperation of all parties. Independent support and evaluation of project. Improved "group intelligence."
a_6		
⋮		
a_{15}		

Now, let us use the idea of quantifying these narrative outcomes through the use of utilities. First, establish which outcome, of the 10 listed, is the more desirable. I would select Action a_5 if θ_2 were the true state of nature. We will give that outcome a utility value of 1.0. Now, let us identify the least desirable outcome. For this I would probably select a_3 if θ_1 <u>or</u> θ_2 were the true state of nature. Let us give this outcome a utility value of 0.0. All of the other outcomes can be thought of as having some weights, or levels of desirability between the best and the worst. I have indicated my own "approximate" values of each of the remaining outcomes. These are shown in Table 2.

	1	2
a_1 ----------	.4	.8
a_2 ----------	.2	.4
a_3 ----------	0.0	0.0
a_4 ----------	.3	.3
a_5 ----------	.6	1.0
'		
'		
'		
a_{15}		

TABLE 2

I think it is apparent that regardless of the selection of the data gathering technique, the state of nature θ_2, dominates θ_1; i.e., θ_2 is at least as desirable as θ_1 for all actions and is more desirable for some actions. This, of course, implies that you agree with my utilities.

To proceed one step further, now consider the interactions of the project personnel and the Risk Analyst to be a two-person game, with actions of each and outcomes displayed in the matrix below, Table 3.

PLAYER A (Project Person)

PLAYER B (Risk Analyst)

	Freudian Approach	Rogerian Approach
Freudian Approach	Both directive. Both defensive. Avoid each other. Contacts - Conflict R.A. turns to modelism	Project personnel attempt to cooperate. R.A. tends to turn project person away. R.A. threatening to project person.
Rogerian Approach	Initially project feels threatened by R.A. Cooperation may later develop. At least project person does not resent R.A.'s involvement.	Harmony. No conflict. Both concur in value and conclusions of the risk analysis.

TABLE 3

The narrative statements of the outcomes in each box of the matrix are those applied to the organization of which Players A and B are members. Once again, using my utilities, the matrix becomes:

PLAYER A

PLAYER B

	Freudian Approach	Rogerian Approach
Freudian Approach	.1	.3
Rogerian Approach	.7	1.0

TABLE 4

Once again, the optimal strategy for the Risk Analyst is to adopt a Rogerian approach, regardless of the attitude of the Project Person, such that the utility to the organization will be greatest.

Robert Boguslow, in his book, "The New Utopians", characterizes a new breed of scientists, called, "Operations Research Analysts," as the New Utopians. Boguslow asserts that these New Utopians are trying to achieve a Utopian Society free from human imperfection. Unlike the classical utopians, such as Thomas Moore and Frances Bacon, who were concerned with solutions to problems as a means of improving the welfare of human beings, Boguslow claims that the New Utopians are preoccupied only with efficiency and people substitutes. They seek Utopia, he claims, by addressing solutions which call for decreases in the number and responsibility of human beings. He suggests that problems confronting these O.R. analysts appear to them as an iceberg where only perhaps one-tenth of the total iceberg is visible above the surface of the water, and he chides them for not attempting to uncover the bottom nine-tenths of the iceberg wherein lie all of the subtle interactive, dysfunctional consequences of the actions dictated by the visible portion of the iceberg.

I have noticed a stigma attached to the entire field of O.R./Risk/Systems Analysis as typified by Boguslow's comments. I feel it is now time for the O.R./Risk Analyst to come down out of his Ivory Tower and get into the foxholes. It is my contention that when he does, as I believe he must, he will be far more successful if he practices the approaches of Client-Centered Risk Analysis. In doing so he must be particularly en-guarde against confusing a client-centered approach for a complacent, laissez-faire approach and he must be equally en-guarde against being co-opted by the project personnel. The tasks for the Risk Analyst are not easy and the tradeoffs may be somewhat uncomfortable and, in some situations, they may even be impossible.

I believe it was Plato who said he would ban poets from his Utopia, as poets have an uncanny knack for perceiving and describing the truth. Analogously, the Risk Analyst must be careful so as not to be banned or circumvented by his organization. His role can be great; his contributions significant.

At the outset of one of the greatest conflicts in the history of mankind, W.H. Auden, the poet, wrote in his poem titled, "September 1, 1939,"

> Into this neutral air
> Where blind skyscrapers use
> Their full height to proclaim
> The strength of Collective Man
> Each language pours its vain
> Competitive excuse:

I believe the Risk Analyst should give the client-centered approach a try. The only danger, is that it may spread to other parts of the organization.

END